Achievement
Testing
Program

Provincial Report
June 1989 Administration

Students First !

Student Evaluation and Records

Alberta
EDUCATION

September 1989

90-02579/

DISTRIBUTION: Superintendents of Schools • School Principals and Teachers • The Alberta Teachers' Association • Alberta School Trustees' Association • Alberta Education • General Public upon Request

PROVINCIAL REPORT
JUNE 1989 ADMINISTRATION

MESSAGE FROM THE DIRECTOR

On behalf of the Student Evaluation and Records Branch, I am pleased to present the Achievement Testing Program *Provincial Report*.

In June 1989, achievement tests were administered in Grade 3 English Language Arts, Grade 6 Social Studies, and Grade 9 Science. Results show that overall student performance remains satisfactory. More students than expected in Grade 3 Language Arts achieved the acceptable standard but fewer students than expected in Grade 6 Social Studies and in Grade 9 Science achieved the acceptable standard. The number of students achieving excellence exceeded expectations in all three testing areas. This provincial pattern of results is consistent with previous test administrations. I particularly draw readers' attention to Section 7, where 1989 results are compared with those obtained in 1985.

This year, more Francophone and French Immersion students than ever before participated in our testing program. For the first time, we are providing jurisdictions with their French language program results in relation to provincial data in order to open discussions on the most useful way of reporting Francophone and French Immersion results. We also carried out special studies to find out the effects of language of testing on measuring achievement in the French Immersion Program. Results for French language testing will be available to administrators and teachers in participating jurisdictions and schools in the form of a special report. Other interested individuals may receive a copy of this special report upon request.

Achievement test results provide educators with an opportunity to inform parents about the strengths of local programs and about initiatives taken to address any weaknesses. We have included in Appendix F the answers to a number of questions frequently posed by parents concerning the Achievement Testing Program. Local boards, schools, and teachers may find it useful to reproduce and distribute these pages to interested parents or to use them as an item in district and school newsletters.

The staff at the Student Evaluation and Records Branch have tried to make the *Provincial Report* more readable and informative. We hope you like these improvements. Please let us know your opinion by completing the questionnaire at the end of this document.

Finally, I wish to express special appreciation to those teachers, principals, and superintendents who shared their experience and expertise with the Student Evaluation and Records Branch during the rigorous achievement test development and marking processes. Your dedication to those tasks and to making certain that test conditions were uniform has once again enabled us to assess student achievement fairly and accurately across Alberta. We are pleased to have your assistance in providing valuable information about student achievement to policymakers, educators, and the public.

Frank G. Horvath, Director

Students First !

Student Evaluation and Records

CONTENTS

	PAGE
Section 1: Summary of Achievement Test Results	1
Section 2: Guidelines for Interpreting Achievement Test Results	3
Section 3: Grade 3 English Language Arts	7
Section 4: Grade 6 Social Studies	21
Section 5: Grade 9 Science	35
Section 6: Student Achievement By Gender	45
Section 7: Student Achievement Over Time	49
Appendix A: Using Achievement Test Results	57
Appendix B: Developing Achievement Tests	59
Appendix C: Standard Setting	63
Appendix D: Equating Multiple-Choice Test Scores	64
Appendix E: Equating Written-Response Test Scores	65
Appendix F: Reporting to Parents: Answers to Frequently Asked Questions	67
Appendix G: Results in Relation to Standards	69
Achievement Testing Program: *Provincial Report* Questionnaire	end page

LIST OF FIGURES

FIGURE PAGE

SECTION 1

1-1 Percentage of Population Writing Achievement Tests June 1989 1

1-2 Percentage of Students Achieving Acceptable Standard June 1989 2

1-3 Percentage of Students Achieving Standard of Excellence June 1989 2

SECTION 6

6-1 Number of Tests Written, By Gender, for Each Subject 45

6-2 Percentage of Students Achieving Acceptable Standard By Gender, for Each Subject 46

6-3 Percentage of Students Achieving Standard of Excellence By Gender, for Each Subject 46

APPENDIX G

G-1 Grade 3 English Language Arts: Percentage of Students Achieving the Acceptable Standard Based on the English Language Program Students Who Wrote the Test June 1989 69

G-2 Grade 3 English Language Arts: Percentage of Students Achieving the Acceptable Standard Based on All Students in the English Language Program June 1989 70

G-3 Grade 6 Social Studies: Percentage of Students Achieving the Acceptable Standard Based on the English Language Program Students Who Wrote the Test June 1989 70

FIGURE PAGE

APPENDIX G (continued)

G-4 Grade 6 Social Studies: Percentage of Students Achieving the Acceptable Standard Based on All Students in the English Language Program June 1989 71

G-5 Grade 9 Science: Percentage of Students Achieving the Acceptable Standard Based on the English Language Program Students Who Wrote the Test June 1989 71

G-6 Grade 9 Science: Percentage of Students Achieving the Acceptable Standard Based on All Students in the English Language Program June 1989 72

LIST OF TABLES

TABLE PAGE

SECTION 3

3-1 Grade 3 English Language Arts: Part B: Reading Achievement Test Blueprint 8

3-2 Grade 3 English Language Arts: Students Tested 9

3-3 Grade 3 English Language Arts: Students Included in Provincial Results, Absentees, and Exemptions 9

3-4 Grade 3 English Language Arts: Percentage of Students Achieving Standards 10

3-5 Grade 3 English Language Arts: Percentage Distribution of Jurisdictions Meeting Expectations 11

3-6 Grade 3 English Language Arts: Frequency Distribution of Total Test Scores 12

TABLE PAGE

SECTION 3 (continued)

3-7 Grade 3 English Language Arts: Part A: Writing Frequency Distribution of Raw Scores 13

3-8 Grade 3 English Language Arts: Part A: Writing Percentage Distribution of Scores 14

3-9 Grade 3 English Language Arts: Part B: Reading Raw Score Results by Reporting Category 15

3-10 Grade 3 English Language Arts: Part B: Reading Frequency Distribution of Raw Scores 16

3-11 Grade 3 English Language Arts: Part B: Reading Results for Individual Multiple-Choice Questions 17

SECTION 4

4-1 Grade 6 Social Studies: Part A: Multiple Choice Achievement Test Blueprint 22

4-2 Grade 6 Social Studies: Part B: Written Response Achievement Test Blueprint 23

4-3 Grade 6 Social Studies: Students Tested 24

4-4 Grade 6 Social Studies: Students Included in Provincial Results, Absentees, and Exemptions 25

4-5 Grade 6 Social Studies: Percentage of Students Achieving Standards 25

4-6 Grade 6 Social Studies: Percentage Distribution of Jurisdictions Meeting Expectations 26

4-7 Grade 6 Social Studies: Frequency Distribution of Total Test Scores 27

4-8 Grade 6 Social Studies: Part A: Multiple Choice Raw Score Results by Reporting Category 29

TABLE PAGE

SECTION 4 (continued)

4-9 Grade 6 Social Studies: Part A: Multiple Choice Frequency Distribution of Raw Scores 30

4-10 Grade 6 Social Studies: Part A: Multiple Choice Results for Individual Multiple-Choice Questions 31

4-11 Grade 6 Social Studies: Part B: Written Response Average Scores Awarded 32

4-12 Grade 6 Social Studies: Part B: Written Response Distribution of Scores for Short-answer Questions 33

4-13 Grade 6 Social Studies: Part B: Written Response Distribution of Scores for Composition Question 33

SECTION 5

5-1 Grade 9 Science: Achievement Test Blueprint 36

5-2 Grade 9 Science: Students Tested 37

5-3 Grade 9 Science: Students Included in Provincial Results, Absentees, and Exemptions 38

5-4 Grade 9 Science: Percentage of Students Achieving Standards 38

5-5 Grade 9 Science: Percentage Distribution of Jurisdictions Meeting Expectations 39

5-6 Grade 9 Science: Raw Score Results by Reporting Categories 40

5-7 Grade 9 Science: Frequency Distribution of Raw Scores 41

5-8 Grade 9 Science: Results for Individual Multiple-Choice Questions . . . 42

TABLE PAGE

SECTION 6

6-1 Number and Percentage of Students Writing Achievement Tests by Gender 47

6-2 Grades 3, 6, and 9 Achievement Tests Average and Standard Deviation by Gender 47

SECTION 7

7-1 Grade 3 English Language Arts: Comparison of 1985 and 1989 Test Results 50

7-2 Grade 3 English Language Arts: Comparison of Percentage of Students Achieving Standards 1985 and 1989 51

7-3 Grade 6 Social Studies: Comparison of 1985 and 1989 Test Results . . . 52

7-4 Grade 6 Social Studies: Comparison of Percentage of Students Achieving Standards 1985 and 1989 53

7-5 Grade 9 Science: Comparison of 1985 and 1989 Test Results 54

7-6 Grade 9 Science: Comparison of Percentage of Students Achieving Standards 1985 and 1989 55

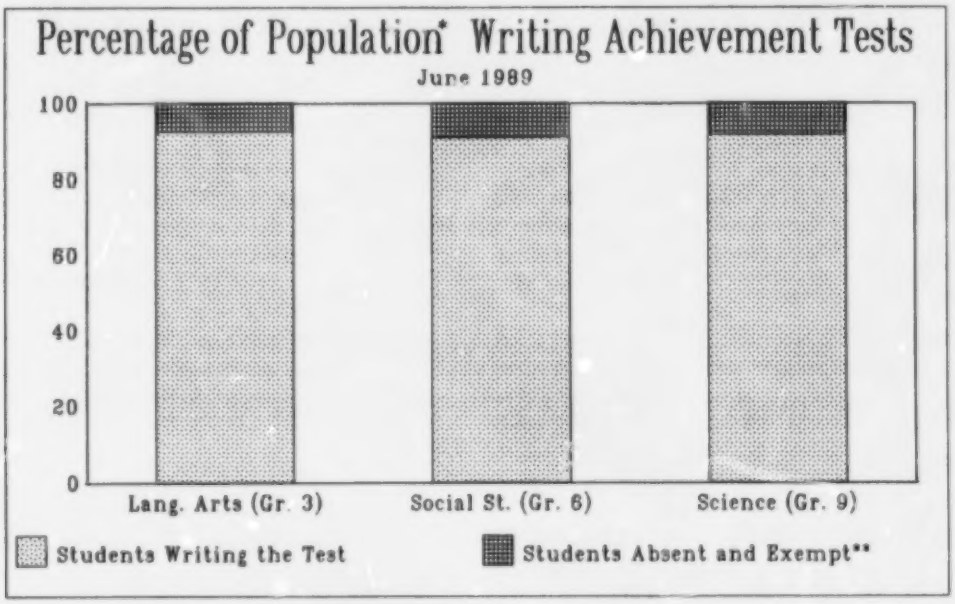
SECTION 1
SUMMARY OF ACHIEVEMENT TEST RESULTS

STUDENT POPULATION

In June 1989, 91 286 achievement tests were administered to students in the province of Alberta.

Figure 1-1 shows the percentage of students in grades 3, 6, and 9 who wrote the achievement tests as well as those who were absent and exempt.

Figure 1-1



* Students in the English language program who were required to participate in the Achievement Testing Program.

** See tables 3-3, 4-4, and 5-3.

The number of students who were absent on the day the tests were administered or who were exempt from writing is shown in tables 3-3, 4-4, and 5-3.

RESULTS IN RELATION TO STANDARDS

Through discussions with educators and school administrators and from our experience with measuring student achievement according to the expectations in the Program of Studies, we believe that it is reasonable to expect that at least 85% of students should achieve at the acceptable level and at least 15% of students should achieve at the level of excellence. As in previous years, results are reported in relation to these standards. Standards are based only on results achieved by English language program students who wrote the 1989 achievement tests and not on the total population. Results in relation to the acceptable standard are presented for the total population in Appendix G, page 69. The total population comprises all English language program students who wrote the test and those students who were absent or who were exempt from writing the test.

Figures 1-2 and 1-3, page 2, present the percentage of students who met the acceptable standard and the standard of excellence.

Figure 1-2

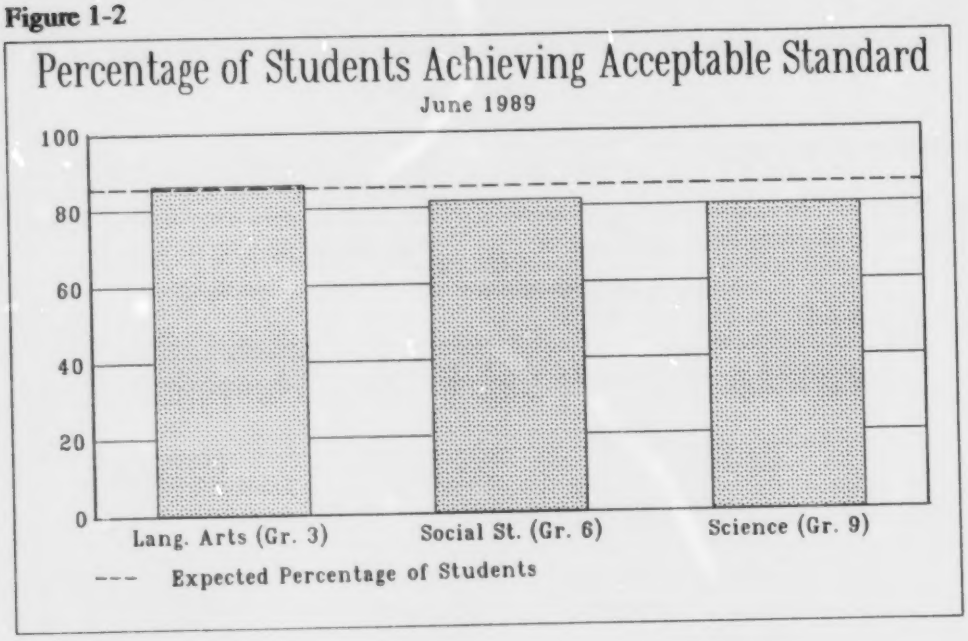
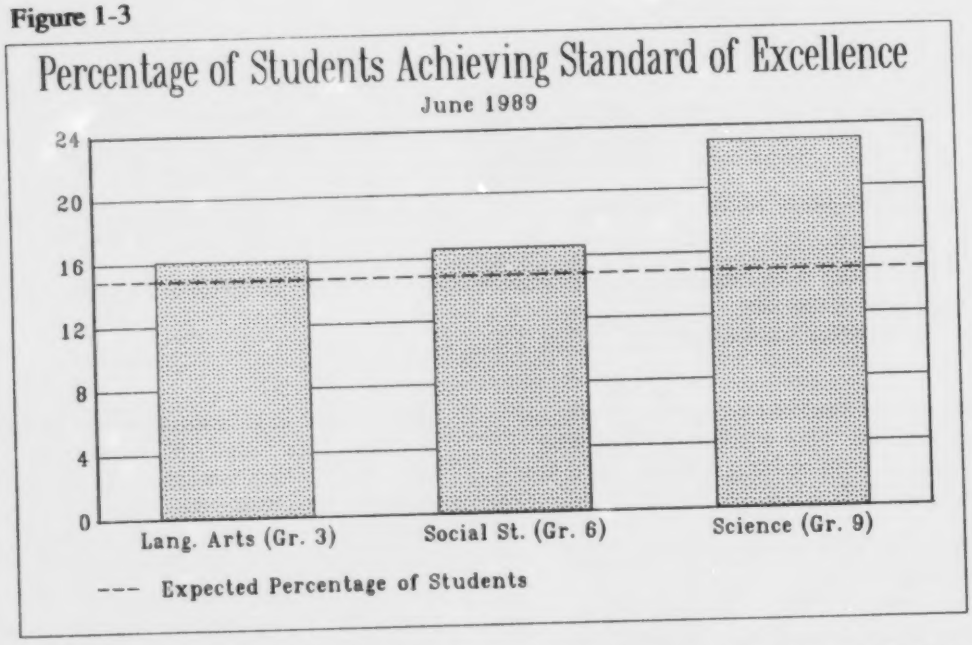


Figure 1-3



Results presented in figures 1-2 and 1-3 reveal that the percentage of students achieving the acceptable standard was slightly higher than was expected for Grade 3 but the respective percentages were lower than were expected for grades 6 and 9. For the standard of excellence, percentages were slightly higher than were expected for grades 3 and 6 and were considerably higher than were expected for Grade 9.

SECTION 2
GUIDELINES FOR INTERPRETING ACHIEVEMENT TEST RESULTS

Following each administration of the achievement tests, a *Provincial Report* is prepared. This report is a public document that describes the aggregated results obtained by those students who wrote achievement tests in a given year. Provincial reports can be used by school board members, superintendents, principals, and teachers as they review their own confidential jurisdiction and school reports.

By using the *Provincial Report* in this way, policymakers and educators can check their perceptions of local achievement against provincewide standards and trends in the levels of achievement.

This *Provincial Report* describes the results achieved by students who wrote the June 1989 achievement tests in Grade 3 English Language Arts, Grade 6 Social Studies, and Grade 9 Science.

The achievement test development process is described in Appendix B, page 59.

Provincial results for students who wrote French translations of the Grade 6 Social Studies Achievement Test and the Grade 9 Science Achievement Test will be presented in a special report. This report will be available to participating jurisdictions and schools. Other interested individuals may receive a copy of this special report upon request.

JURISDICTION AND SCHOOL REPORTS

In addition to the *Provincial Report*, superintendents and principals receive confidential reports of results achieved by students in their particular jurisdiction or school. The jurisdiction and school reports contain tables that parallel the major tables in the *Provincial Report*.

Policymakers and educators in each jurisdiction are encouraged to study carefully the provincial results and their own test results.

Educators at the school and jurisdiction level can make two kinds of comparisons of the achievement of their students. One comparison is in relation to expectations or standards; the other is in relation to the achievement of students in the entire province of Alberta.

As a result of these comparisons, teachers, principals, and superintendents can reflect on the programs that were delivered in their grades 3, 6, and 9 classrooms and make changes wherever necessary or desirable.

USE OF THE JURISDICTION AND SCHOOL REPORTS

The reports are NOT intended to be used as the basis for

- making decisions about student placement or promotion
- evaluating teacher performance
- comparing performance between or among schools.

Administrators in each jurisdiction should apply separate locally developed teacher, school, and school system evaluation policies to the tasks of evaluating teacher and school performance.

The information provided in the reports is factual regarding *what* has happened as a result of the administration of the tests. The interpretation of this information -- hypothesizing *why* results are as they are -- involves consideration of the many factors and variables that contribute to achievement.

In addition, it must be noted that the information in these reports is itself limited to selected objectives of the Program of Studies. Many important aspects of learning cannot be measured by the achievement tests, which are time-limited paper and pencil tests.

STANDARD SETTING

Standards have been set for each of the achievement tests (see Appendix C, page 63). Our judgment is that 85% of students should be able to meet or exceed the acceptable standard of achievement, and 15% should be able to meet or exceed the standard of excellence. Included in the *jurisdiction* and *school reports* is a table showing the percentage of students meeting each standard set. The table also indicates whether the number of students in that school or jurisdiction who have achieved the standard is significantly different from the expected number (based on the 85% and 15% expectations), and unlikely to be due to chance variation.

For the purposes of the Achievement Testing Program, the 95% confidence interval is used. That is, if the probability is less than one in 20 that a difference is due to chance, this difference is very likely a real difference.

Although the statistical tests take the number of students into consideration, it is a useful rule of thumb that results for groups of fewer that 25 students must be interpreted with particular caution. Chance variation in small groups is greater.

Educators interpreting these reports are encouraged to consider how well their students have done compared to the standards.

COMPARING RESULTS TO AVERAGE SCORES

While overall test results are presented in relation to provincial standards, each jurisdiction and school report also provides jurisdiction or school average scores for each reporting category or subtest. Each of these scores may be compared to the provincial average for the same reporting category or subtest to determine if differences exist.

The importance of differences that may exist between jurisdiction or school averages and provincial averages is not always clear. To aid in the interpretation of differences between the averages, jurisdiction and school reports indicate when a difference is unlikely to be due to chance variation. The 95% confidence interval is also used here to identify significant differences.

FACTORS LIMITING THE INTERPRETATION OF TEST RESULTS

Educators who are interpreting results must take into account the following limitations:

1. Paper and pencil tests necessarily measure reading ability. Achievement tests are designed to have a readability level equivalent to the grade level being tested. Jurisdictions should consider the average reading level of their grades 6 and 9 students, as reading levels below these grades will have an effect on test results that will be independent of achievement in social studies and science respectively.

2. If more than 10% of eligible students in a jurisdiction did not write a test, the reported averages for that jurisdiction may not accurately represent the true averages.

3. Consideration should be given to the degree to which students in particular classes or grades were motivated to perform to their levels of ability.

FACTORS THAT MAY AFFECT STUDENT ACHIEVEMENT

There are many factors or variables that may effect student achievement. Some of these factors are:

1. Environment

   - community environment
   - school environment
   - socioeconomic background
   - family circumstances

2. Student Factors

   - ability
   - attitude
   - motivation
   - aspiration
   - academic background
   - learning style

3. Resources (availability and appropriateness)

   - programs of study
   - curriculum of study
   - resource materials
   - library services
   - current textbooks
   - references

4. Instruction

   - teacher qualifications
   - teacher experience
   - professional development
   - teacher morale
   - teaching strategies
   - hours of instruction
   - staff turnover
   - amount of homework assigned
   - communication of teacher expectations

SECTION 3
GRADE 3 ENGLISH LANGUAGE ARTS

GENERAL DESCRIPTION

The Grade 3 English Language Arts Achievement Test was a two-part test. Part A: Writing was a 60-minute writing test consisting of a story starter and instructions for the student to finish writing the story. This format was designed to reflect the writing process. Part B: Reading was a 50-minute reading test consisting of 40 multiple-choice questions based on reading selections from fiction, nonfiction, and poetry.

The test was designed to reflect the Grade 3 Language Arts curriculum specifications that have been developed from the *Program of Studies for Elementary School 1978* (amended 1982). The scope of the Grade 3 English Language Arts Achievement Test was limited to the writing and the reading components of the program.

The information presented in this section is based on the results achieved by 31 998 students.

SUMMARY OF RESULTS

Results in Relation to Standards

Results show that 86.2% of students who wrote the test achieved the acceptable standard and 16.1% achieved the standard of excellence. These results were slightly higher than expectations for both standards. The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Average Score

The average total score for the test was 68.9%, with a standard deviation of 15.7. The average raw score for Part A: Writing was 16.2 marks out of a possible 25, with a standard deviation of 4.1. For Part B: Reading, the average raw score was 29.1 marks out of a possible 40, with a standard deviation of 8.1.

CONTENT OF PART A: WRITING

The Part A: Writing booklet included one page labelled IDEAS and several pages for completing the story. The writing assignment followed a story starter that was read by the student. The assignment set a specific writing task but allowed the student to use imagination and background experience to develop a story. Papers were scored for Content, Development, Sentence Structure, Vocabulary, and Conventions.

CONTENT OF PART B: READING

Every effort was made to select complete passages for Part B: Reading. As well, reading selections were chosen to reflect the interests of the majority of Grade 3 students and to be of appropriate difficulty for Grade 3 students. Extensive use was made of Canadian material.

Questions were developed to test how well the students could understand and analyse the reading selections, and could make judgments about form and content. Only questions dealing with significant aspects of the reading selections were used.

ACHIEVEMENT TEST BLUEPRINT

Questions were classified according to two cognitive levels: Literal Understanding (12 questions), and Inferential Understanding and Judgment (28 questions). By considering cognitive level when developing a test, the Student Evaluation and Records Branch ensures that students will use a variety of mental activities as they write the test.

Questions listed under Literal Understanding are designed to test the skills of recall and recognition; those listed under Inferential Understanding and Judgment are designed to test the skills of analysis, interpretation, extrapolation, and judgment.

Table 3-1 presents the blueprint used to develop Part B: Reading. Classification by reporting category for each question included in Part B: Reading is indicated in the table.

Table 3-1
Grade 3 English Language Arts
Part B: Reading
Achievement Test Blueprint

Reporting Category	Cognitive Level		Total Number of Questions
	Literal Understanding	Inferential Understanding and Judgment	
1. Attending to Details The student should be able to construct meaning from background experience and by attending to the supporting details found in a reading selection.	*1,5,13,16, 19,24,26, 29,30,31*	*2,9,14,17,18, 23,25,27,28, 32,33,37*	22
2. Associating Meanings The student should be able to associate meanings of words and expressions from background experience and from contextual clues in a reading selection.	*11,12*	*4,7,8,10,22, 40*	8
3. Synthesizing Ideas The student should be able to synthesize ideas from the entire reading selection in order to construct meaning, to deduce the main idea, and to predict plausible outcomes or conclusions.		*3,6,15,20,21, 34,35,36,38, 39*	10
Total	12	28	40

STUDENTS TESTED, ABSENT, AND EXEMPT

Table 3-2 presents the number of students who wrote the Grade 3 English Language Arts Achievement Test. Students in French Immersion or Francophone programs could be exempt from the test at the option of the superintendent. As the table shows, 2 092 of these students wrote the test. Their scores are NOT included in the results given in this *Provincial Report* but will be provided in a special report, which will be available to administrators and teachers in participating jurisdictions and schools. Another 160 students in French Immersion or Francophone programs did not write the test, as Table 3-3 indicates.

Table 3-2
Grade 3 English Language Arts
Students Tested

Type of Participation	Number of Students
Participation Required (Students in Regular Program)	31 998*
Participation Optional (Students in Francophone/French Immersion Programs)	2 092**

* Of the total number of students required to write the test, 1 146 students were absent the day the test was written and 1 500 students were exempt from writing the test. (See Table 3-3.)

** Results achieved by these students are not included in the provincial data because participation in the Achievement Testing Program is optional for these students

Table 3-3 presents the number and percentage distribution of students who were required to write the Grade 3 English Language Arts Achievement Test and who were absent or exempt.

Table 3-3
Grade 3 English Language Arts
Students Included in Provincial Results, Absentees, and Exemptions

Category	Number of Students	Percentage of Students
Students Included in Provincial Results	31 998	92.4
Students Absent	1 146	3.3
Students Exempt:	1 500	4.3
Special Needs	887	2.6
Not Taught Subject This Term	21	0.1
ESL Classes	408	1.2
French Immersion/Francophone Program	160	0.5
Other (approved by Director)	24	0.1

RESULTS FOR THE TOTAL TEST

To calculate a total test score for students, Part A: Writing and Part B: Reading were given equal weighting. A summary score for Part A: Writing was calculated by adding up the scores for each of the five written-response scales, thus giving each scale equal weighting in the summary score.

RESULTS IN RELATION TO STANDARDS

Table 3-4 shows the percentage of students achieving the acceptable standard and the standard of excellence for the total test, for Part A: Writing, and for Part B: Reading.

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Table 3-4
Grade 3 English Language Arts
Percentage of Students Achieving Standards

Reporting Category and Level of Standard	Score Representing Standard	Percentage of Students Achieving At or Above Standard	
		Expected	Actual
Total Test (Maximum Possible Score = 100)			
Acceptable Standard	51	85.0	86.2
Standard of Excellence	85	15.0	16.1
Part A: Writing (Maximum Possible Raw Score = 25)			
Acceptable Standard	13	85.0	84.5
Standard of Excellence	20	15.0	20.6
Part B: Reading (Maximum Possible Raw Score = 40)			
Acceptable Standard	21	85.0	82.7
Standard of Excellence	36	15.0	25.4

The numbers of students achieving the acceptable standard and the standard of excellence for each jurisdiction were analysed to determine whether jurisdictions were below expectations, meeting expectations, or above expectations. Jurisdictions classified as meeting expectations were those for which the difference between the actual number of students and the expected number of students at or above expectations was not statistically significant. A 95% confidence interval was used; this criterion means that differences are only reported when there is a 5% or smaller probability that a difference of that size could occur by chance.

The results are reported in Table 3-5. The percentages in the table are based on 207 jurisdictions (including private schools).

Table 3-5
Grade 3 English Language Arts
Percentage Distribution of Jurisdictions* Meeting Expectations

Reporting Category and Level of Standard	Below Expectations	Meeting Expectations	Above Expectations
Total Test			
Acceptable Standard**	13.5	64.7	21.7
Standard of Excellence***	5.8	77.8	16.4
Part A: Writing			
Acceptable Standard	12.6	76.8	9.7
Standard of Excellence	1.9	61.4	36.7
Part B: Reading			
Acceptable Standard	17.9	72.8	9.7
Standard of Excellence	1.4	42.5	56.0

* Jurisdictions with fewer than five students are excluded, as the statistical significance of the frequencies compared to the expectations cannot be calculated.
** Acceptable Standard: 85% of students are expected to achieve at or above the acceptable standard.
*** Standard of Excellence: 15% of students are expected to achieve at or above the standard of excellence.

AVERAGE SCORE

Another way to look at the achievement of students is by means of the average score.

The average total score for the Grade 3 English Language Arts Achievement Test was 68.9%, with a standard deviation of 15.7.

Table 3-6 shows the percentage of students who obtained each total test score (relative frequency) and the percentage of students who scored at or below each total test score (cumulative frequency). Total test scores are expressed as percentages.

Table 3-6
Grade 3 English Language Arts
Frequency Distribution of Total Test Scores

Total Score (%)	Relative Frequency (%)	Cumulative Frequency (%)	Total Score (%)	Relative Frequency (%)	Cumulative Frequency (%)
0	0.0	0.0	51	1.1	15.0
1	0.0	0.0	52	1.1	16.1
2	0.0	0.0	53	1.3	17.4
3	0.0	0.0	54	1.5	18.8
4	0.0	0.0	55	1.4	20.2
5	0.0	0.0	56	1.4	21.6
6	0.0	0.0	57	1.4	23.0
7	0.0	0.0	58	1.5	24.6
8	0.0	0.0	59	1.7	26.3
9	0.0	0.1	60	1.6	27.9
10	0.0	0.1	61	1.8	29.8
11	0.0	0.1	62	1.7	31.5
12	0.0	0.1	63	1.9	33.4
13	0.0	0.1	64	2.2	35.5
14	0.0	0.1	65	2.1	37.6
15	0.0	0.1	66	2.2	39.9
16	0.0	0.2	67	2.2	42.1
17	0.0	0.2	68	2.4	44.5
18	0.0	0.2	69	2.7	47.2
19	0.0	0.2	70	2.5	49.7
20	0.1	0.3	71	2.7	52.4
21	0.0	0.3	72	2.4	54.8
22	0.0	0.3	73	2.9	57.7
23	0.1	0.4	74	2.9	60.7
24	0.1	0.5	75	2.8	63.5
25	0.1	0.6	76	2.6	66.1
26	0.1	0.7	77	2.3	68.5
27	0.1	0.7	78	2.6	71.1
28	0.1	0.9	79	2.5	73.6
29	0.2	1.0	80	2.5	76.1
30	0.2	1.2	81	2.0	78.1
31	0.2	1.4	82	2.0	80.1
32	0.2	1.7	83	1.7	81.8
33	0.3	2.0	84	2.1	83.9
34	0.3	2.3	85	1.6	85.4
35	0.4	2.7	86	1.9	87.4
36	0.4	3.1	87	1.2	88.5
37	0.4	3.5	88	1.6	90.1
38	0.5	4.0	89	1.2	91.3
39	0.6	4.6	90	1.3	92.6
40	0.6	5.2	91	0.9	93.5
41	0.6	5.8	92	0.9	94.4
42	0.7	6.4	93	0.8	95.1
43	0.7	7.2	94	1.1	96.3
44	0.8	7.9	95	0.7	97.0
45	0.9	8.9	96	1.0	98.0
46	0.9	9.8	97	0.3	98.3
47	0.8	10.6	98	0.8	99.1
48	1.0	11.6	99	0.6	99.7
49	1.1	12.7	100	0.3	100.0
50	1.1	13.8			

RESULTS FOR PART A: WRITING

Raw scores were calculated by adding the marks earned for each of the five 5-point reporting categories.

RESULTS IN RELATION TO STANDARDS

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63). For Part A: Writing, the standard established was such that in order to meet

- the acceptable standard, students had to achieve a raw score of 13 out of 25
- the standard of excellence, students had to achieve a raw score of 20 out of 25.

Based on these standards, the results revealed that

- 84.5% of students performed at or above the acceptable standard, and
- 20.6% of students performed at or above the standard of excellence.

These levels of performance were as high as were expected at the acceptable standard and were higher than were expected at the standard of excellence.

AVERAGE SCORE

The average raw score for Part A: Writing was 16.2 marks out of a possible 25, with a standard deviation of 4.1.

Table 3-7 shows the percentage of students who obtained each score on Part A: Writing (relative frequency) and the percentage of students who scored at or below each score (cumulative frequency).

Results for Part A: Writing are most clearly understood in the context of the assignment students responded to and in the context of the scoring descriptors. Complete scoring guides are available from the Student Evaluation and Records Branch (427-2948).

All schools should have extra copies of the Part A: Writing test to use in conjunction with information provided in this *Provincial Report*.

Table 3-7
Grade 3 English Language Arts
Part A: Writing
Frequency Distribution of Raw Scores

Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)	Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)
0	0.2	0.2	13	8.6	24.1
1	0.0	0.2	14	9.6	33.7
2	0.0	0.2	15	14.6	48.3
3	0.0	0.2	16	9.1	57.3
4	0.0	0.2	17	9.2	66.6
5	0.5	0.7	18	7.0	73.5
6	0.3	1.0	19	5.9	79.4
7	0.4	1.5	20	5.7	85.1
8	0.8	2.3	21	3.3	88.3
9	1.1	3.3	22	3.0	91.4
10	3.1	6.4	23	2.6	93.9
11	3.5	9.9	24	2.3	96.3
12	5.6	15.5	25	3.7	100.0

SCORING RELIABILITY

Although the papers were scored on a one-marker system, 244 papers were re-marked so that a second set of scores was available for these papers to confirm scoring consistency. Of the scores awarded on the second reading, 90.8% were identical to the original score on the same scale or varied by only one point. It is important to note that the one-marker system produces results that are reliable for groups of 25 or more students. Achievement test scores are not intended to be reliable for individual students.

The results outlined in Table 3-8 are best considered in terms of the percentage of students that markers judged to have presented work that was 3 (Satisfactory) or better for any reporting category.

It is possible to draw conclusions about local program strengths and weaknesses by comparing local percentages of 3 (Satisfactory) or better scores on each reporting category with the provincial averages.

Students do better on some dimensions of the task than on others. (See Examiners' Remarks, page 18.)

Table 3-8
Grade 3 English Language Arts
Part A: Writing
Percentage Distribution of Scores

Score (Scale Points)	Reporting Category				
	Content	Development	Sentence Structure	Vocabulary	Conventions
5 (Excellent)	11.9	10.7	9.8	8.7	11.0
4 (Proficient)	27.3	26.0	24.4	21.5	23.6
3 (Satisfactory)	43.1	40.4	48.7	55.8	43.0
2 (Limited)	15.6	20.4	15.1	12.6	19.6
1 (Poor)	2.0	2.3	1.8	1.2	2.7
INS (Insufficient or No Response)	0.2	0.2	0.2	0.2	0.2

RESULTS FOR PART B: READING

Since over 94.5% of students writing completed Part B: Reading, it was concluded that sufficient time was allotted for writing the test.

RESULTS IN RELATION TO STANDARDS

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63). For Part B: Reading, the standard established was such that in order to meet

- the acceptable standard, students had to achieve a raw score of 21 out of 40
- the standard of excellence, students had to achieve a raw score of 36 out of 40.

Based on these standards, the results revealed that

- 82.7% of students performed at or above the acceptable standard, and
- 25.4% of students performed at or above the standard of excellence.

The level of performance was slightly lower than was expected at the acceptable standard and was much higher than was expected at the standard of excellence.

AVERAGE SCORE

Provincial summary results for Part B: Reading were as follows:

- Provincial Average -- 29.1 marks out of a possible 40
- Standard Deviation -- 8.1

As outlined in the blueprint on page 8, the questions on Part B: Reading were grouped according to reporting categories.

Raw score averages for each of these reporting categories and for Part B: Reading as a whole are presented in Table 3-9. Raw score averages were computed and rounded to one decimal.

Although levels of performance in the different reporting categories appeared to show some variation, caution is advised when comparing them. The sets of questions that made up each category were not selected to be equal in average level of difficulty; therefore, differences may have been due to variations in question difficulty rather than in student performance. The raw score averages can be used, however, in combination with jurisdiction and school results to detect patterns of relative strength or weakness in achievement in each of the categories.

Table 3-9
Grade 3 English Language Arts
Part B: Reading
Raw Score Results by Reporting Category

Reporting Category	Number of Questions	Raw Score Average	Standard Deviation
Total Part B: Reading	40	29.1	8.1
Attending to Details	22	16.6	4.9
Associating Meanings	8	5.7	1.8
Synthesizing Ideas	10	6.8	2.2
Literal Understanding	12	9.3	2.8
Inferential Understanding and Judgment	28	19.8	5.6

Table 3-10 presents the percentage of students who obtained each score on Part B: Reading (relative frequency) and the percentage of students who scored at or below each score (cumulative frequency).

Table 3-10
Grade 3 English Language Arts
Part B: Reading
Frequency Distribution of Raw Scores

Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)	Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)
0	0.0	0.0	21	2.1	19.4
1	0.0	0.1	22	2.4	21.8
2	0.0	0.1	23	2.4	24.1
3	0.0	0.1	24	2.3	26.4
4	0.1	0.2	25	2.6	29.0
5	0.1	0.3	26	2.9	31.9
6	0.1	0.4	27	3.0	34.9
7	0.3	0.7	28	3.5	38.3
8	0.4	1.1	29	3.7	42.1
9	0.6	1.7	30	4.1	46.1
10	0.7	2.4	31	4.7	50.9
11	1.0	3.4	32	5.2	56.1
12	1.2	4.5	33	5.5	61.6
13	1.2	5.7	34	6.3	67.9
14	1.3	7.0	35	6.7	74.6
15	1.5	8.5	36	6.8	81.4
16	1.6	10.1	37	6.9	88.3
17	1.6	11.7	38	6.1	94.4
18	1.8	13.5	39	4.3	98.7
19	1.9	15.3	40	1.3	100.0
20	2.0	17.3			

PERCENTAGE OF STUDENTS CHOOSING EACH ALTERNATIVE

Table 3-11 presents the percentage of students who chose each alternative (A, B, C, and D) for each multiple-choice question on Part B: Reading. The correct response (key) for each question is also identified.

The results shown in Table 3-11 can best be used in conjunction with results presented in jurisdiction and school reports in order to interpret strengths and weaknesses of local programs.

Table 3-11
Grade 3 English Language Arts
Part B: Reading
Results for Individual Multiple-Choice Questions*

Item	Key	Distribution of Responses (%)				Item	Key	Distribution of Responses (%)			
		A	B	C	D			A	B	C	D
1	B	4.1	84.0	5.7	6.0	21	D	16.2	15.2	9.0	58.7
2	A	84.2	8.9	13.7	2.6	22	A	61.6	5.9	21.7	9.6
3	D	12.7	6.5	3.5	76.7	23	C	14.7	8.6	65.1	10.4
4	B	2.8	91.6	2.9	2.2	24	A	83.9	2.9	8.1	3.6
5	B	8.4	86.6	2.8	1.6	25	D	30.5	10.0	8.2	49.4
6	C	4.9	5.7	81.3	7.8	26	A	77.5	6.9	8.4	4.6
7	D	2.4	9.8	15.1	71.7	27	B	5.0	77.4	10.4	4.8
8	C	5.8	16.0	66.1	11.5	28	C	9.1	5.6	62.1	20.4
9	C	7.2	4.9	77.1	10.1	29	B	8.8	67.6	7.9	12.5
10	A	78.5	3.6	8.0	8.8	30	B	18.7	63.7	7.7	7.1
11	D	2.8	2.5	15.6	78.4	31	C	11.8	4.8	75.4	4.8
12	D	5.0	3.9	6.5	83.8	32	A	75.0	7.7	6.5	7.4
13	A	75.6	8.9	9.9	4.5	33	D	7.1	10.1	7.3	71.9
14	D	3.6	4.7	3.4	87.3	34	B	15.5	64.0	11.8	4.5
15	A	79.8	6.3	4.9	7.8	35	A	66.5	14.1	7.4	7.5
16	D	10.4	5.5	6.0	77.2	36	C	5.0	5.9	78.6	6.1
17	C	3.3	4.3	84.7	6.7	37	C	2.3	4.5	86.5	2.0
18	B	7.1	75.6	6.1	10.1	38	B	64.3	20.6	4.7	5.4
19	B	5.9	78.3	4.1	10.3	39	D	5.5	3.0	4.6	81.7
20	C	9.4	9.7	76.3	3.0	40	C	38.5	7.8	37.4	10.9

*The sum of the percentages for each question may be less than 100% because the No Response category is not included. The No Response category does not exceed 5.4% for any one of these questions.

GRADE 3 ENGLISH LANGUAGE ARTS
EXAMINERS' REMARKS

Teacher-markers and standard-setters felt that the Grade 3 English Language Arts Achievement Test reflected the specifications of the curriculum, with one caution: teachers, during instruction, look at learning in a more holistic and developmental way. Every effort was made to address this perspective in developing the writing and reading components of this test. This test is a valid measure of student achievement in these two areas.

PART A: WRITING

Markers were very pleased with the overall quality of students' writing. Students handled the narrative form very well; they created stories that reflected personal experience and ideas from literature and the media. They took risks with such elements as vocabulary, details, sentence structure, and dialogue. Even most of the weaker students expressed ideas quite clearly -- which supports teacher comments that there is more writing going on in classrooms.

The percentage of students who scored 3 (Satisfactory) or better in 1989 was compared with the 1985 figures:

	1989	1985
Content	82.3%	77.5%
Development	77.1%	71.3%
Sentence Structure	82.9%	77.7%
Vocabulary	86.0%	79.5%
Conventions	77.5%	78.3%

Scores in 1989 increased on all marking scales except for Conventions. Only 0.2% of students, compared to 0.4% in 1985, produced written work that was considered to be Insufficient for scoring purposes.

In summary, at least 77.1% of students scored 3 (Satisfactory) or better on any marking scale. The highest achievement was on Vocabulary and the lowest was on Development.

A booklet to provide Grade 3 teachers, administrators, and students with samples of students' writing that exemplify the criteria used to score written responses on the June 1989 Grade 3 English Language Arts Achievement Test will be published in the near future and mailed to teachers.

PART B: READING

Teacher-markers and standard-setters felt that the reading section represented an appropriate range of difficulty for Grade 3 students. They appreciated the variety of themes, the quality literature, and the Canadian content.

Detailed statistical review of all questions revealed that students scored higher on questions that required Literal Understanding -- 77.5% -- than on questions that required Inferential Understanding and Judgment -- 70.7%.

Questions ranged from those that students found to be difficult, such as question 38 with only 20.6% of students answering correctly, to those that were easy for students, such as question 4 with 91.6% of students answering correctly.

Question 38 proved to be difficult because it required a high level of thinking. The question read:

38. Which sentence BEST tells us that the Rabbit was no longer a toy?
    - "'You will be REAL!'"
    - "Love stirred in his little sawdust heart."
    - "The Rabbit was wet through with the dew."
    - "The Boy and his Rabbit had long days in the garden."

It required students to recognize the significance of the difference in tense between the stem and the first alternative. In addition, students had to be aware that love cannot be experienced by an inanimate object such as a toy.

Question 4 asked students to infer the meaning of a word from context. The question read:

4. In the story, what does the underlined word enormous mean?
   - White
   - Large
   - Lonely
   - Straight

Students found this to be a very easy task either because they are familiar with the word "enormous" or because they have a high level of skill in deriving meaning from context.

SECTION 4
GRADE 6 SOCIAL STUDIES

GENERAL DESCRIPTION

The Grade 6 Social Studies Achievement Test was a two-part test. The time allotted for writing each part was 50 minutes.

Part A: Multiple Choice consisted of 50 questions worth 70% of the total test score.

Part B: Written Response consisted of six short-answer questions and a composition question worth 30% of the total test score. All questions in this second part reflected a current social issue.

The information presented in this section is based on the results achieved by 29 918 students.

SUMMARY OF RESULTS

Results in Relation to Standards

Results show that 81.6% of students who wrote the test achieved the acceptable standard and 16.6% achieved the standard of excellence. These results were lower than were expected for the acceptable standard and were slightly higher than were expected for the standard of excellence. The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Average Score

The total test score was obtained by combining the scores for Part A: Multiple Choice and Part B: Written Response so that the two parts had a weighting of 70% and 30% respectively.

The average total score for the test was 62.5%, with a standard deviation of 16.3. The average raw score for Part A: Multiple Choice was 32.2 marks out of a possible 50, with a standard deviation of 8.9. For Part B: Written Response, the average raw score was 17.5 marks out of a possible 30, with a standard deviation of 5.1.

CONTENT OF THE TEST

The Grade 6 Social Studies Achievement Test was based on the *1981 Alberta Social Studies Curriculum*. All test questions were drawn from the content of the three topics prescribed for Grade 6:

- Topic A: How People in Earlier Times Met Their Needs
- Topic B: How People in Eastern Societies Meet Their Needs Today
- Topic C: Meeting Needs Through Local, Provincial, and Federal Governments

Content emphases were drawn from the *Grade 6 Social Studies Curriculum Specifications*.

The Grade 6 Social Studies Achievement Test measured value, knowledge, and skill objectives. Objectives related to the development of attitudes and participation skills did not form part of this test. The weighting allocated to the development of these objectives in the curriculum specifications was reassigned to the remaining objectives on a prorated basis.

PART A: MULTIPLE CHOICE BLUEPRINT

Table 4-1, page 22, presents the blueprint used to develop the Part A: Multiple Choice section of the test. Classification by reporting category for each question on the test is indicated in the table.

Table 4-1
Grade 6 Social Studies
Part A: Multiple Choice
Achievement Test Blueprint

Process Reporting Category	Concept Reporting Category				Percent of Total Score
	Value Concepts Knowledge of competing values or value positions	How People in Earlier Times Met Their Needs Knowledge of facts, concepts, and generalizations related to the meeting of needs in earlier times	How People in Eastern Societies Meet Their Needs Today Knowledge of facts, concepts, and generalizations related to the meeting of needs in eastern societies	Meeting Needs Through Local, Provincial, and Federal Governments Knowledge of facts, concepts, and generalizations related to the meeting of needs through governments	
Knowledge and Comprehension Recalls or recognizes data and transforms data into other words	*3,35*	*5,6,7,8,9, 10,11*	*18,19,20,21, 22,26,33*	*43,44,45,46,47,48, 49,50*	34
Inquiry Skills I Uses skills to identify an issue, select appropriate research questions, and gather and organize data		*1,2,16*	*23,30*	*34,36,37*	11
Inquiry Skills II Uses skills to analyse, evaluate, and synthesize data		*12,13*	*27,28,29,32*	*40,42*	11
Inquiry Skills III Uses skills to resolve an issue, apply a decision, and evaluate that decision		*14,15,17*	*25*	*38,39*	8
Valuing Skills Uses skills to analyse competing values	*4,24,31,41*				6
Percent of Total Score	8	21	20	21	70

PART B: WRITTEN-RESPONSE BLUEPRINT

The written-response section of the test consisted of seven questions. Questions 1 to 6 required short-answer responses.

Question 7 asked students to write two or more paragraphs to persuade their classmates to adopt their position on an issue. The objectives on which the written-response questions were based are shown in Table 4-2.

Table 4-2
Grade 6 Social Studies
Part B: Written Response
Achievement Test Blueprint

Reporting Category	Description of Writing Assignment	Proportion of Total Score (%)
I. Short Answer (Identification of the Elements of an Issue)	1. Recalls facts related to an issue. Knowledge objectives -- recall knowledge.	2
	2. Recalls facts and applies them in a new situation. Skill objectives -- analyse and evaluate data.	2
	3. Recalls facts and applies them in a new situation. Skill objectives -- analyse and evaluate data.	2
	4. Formulates a generalization. Skill objectives -- synthesize data.	2
	5. Identifies speakers' value positions. Value objectives -- develop an understanding of values and analyse values.	3
	6. Identifies speakers' value positions. Value objectives -- develop an understanding of values and analyse values.	4
Subtotal		15
II. Composition (Resolution of an Issue)	7. Presents and defends a position. Skill objectives -- resolve the issue and communicate effectively.	15
Subtotal		15
Total		30

STUDENTS TESTED, ABSENT, AND EXEMPT

Table 4-3 presents the number of students who wrote the Grade 6 Social Studies Achievement Test or its French translation (6^{e} Année Test de Rendement Etudes sociales). Students in French Immersion or Francophone programs could be exempt from the test at the option of the superintendent. As the table shows, 77 of these students wrote the test in English, and 1 104 wrote the test in French translation. Because their participation was optional, their scores are NOT included in the results given in this section of the *Provincial Report*. Results for students in French Immersion or Francophone programs who wrote the French translation will be presented in a special report, which will be available to administrators and teachers in participating jurisdictions and schools. Of the 173 students listed in Table 4-4 as exempt because the language of instruction was not English, an undetermined number were in French Immersion or Francophone programs.

A special study to determine the effects of the language of testing on achievement scores was conducted by administering the Grade 6 Social Studies Achievement Test and its French translation to students in selected French Immersion Program classes. Results for students participating in this study will be provided in a special report under separate cover.

Table 4-3
Grade 6 Social Studies
Students Tested

Type of Participation	Number of Students
Participation Required (Students Receiving Instruction in English)	29 918*
Participation Optional (Students Receiving Instruction in French)	
Wrote in English	77**
Wrote French Translation	1 104***
Selected Participation in Special Study (Students in French Immersion Program)	
Wrote in English	216***
Wrote French Translation	221***

* Of the total number of students required to write the test, 1 029 students were absent the day the test was written and 1 970 students were exempt from writing the test. (See Table 4-4.)

** Results achieved by these students are not included in the provincial data because participation in the Achievement Testing Program is optional for these students.

*** Results achieved by these students will be presented in a special report.

Table 4-4 presents the number and percentage distribution of students who were required to write the Grade 6 Social Studies Achievement Test and who were absent or exempt.

Table 4-4
Grade 6 Social Studies
Students Included in Provincial Results, Absentees, and Exemptions

Category	Number of Students	Percentage of Students
Students Included in Provincial Results	29 918	90.9
Students Absent	1 029	3.1
Students Exempt:	1 970	6.0
Special Needs	746	2.3
Not Taught Subject This Term	747	2.3
ESL Classes	276	0.8
Language of Instruction Not English	173	0.5
Other (approved by Director)	28	0.1

RESULTS FOR THE TOTAL TEST

RESULTS IN RELATION TO STANDARDS

Table 4-5 shows the percentage of students achieving the acceptable standard and the standard of excellence for the total test, for Part A: Multiple Choice, and for Part B: Written Response.

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Table 4-5
Grade 6 Social Studies
Percentage of Students Achieving Standards

Reporting Category and Level of Standard	Score Representing Standard	Percentage of Students Achieving At or Above Standard	
		Expected	Actual
Total Test (Maximum Possible Score = 100)			
Acceptable Standard	47	85.0	81.6
Standard of Excellence	80	15.0	16.6
Part A: Multiple Choice (Maximum Possible Raw Score = 50)			
Acceptable Standard	23	85.0	83.3
Standard of Excellence	40	15.0	24.3
Part B: Written Response (Maximum Possible Raw Score = 30)			
Acceptable Standard	15	85.0	72.9
Standard of Excellence	24	15.0	11.9

The numbers of students achieving the acceptable standard and the standard of excellence for each jurisdiction were analysed to determine whether jurisdictions were below expectations, meeting expectations, or above expectations. Jurisdictions classified as meeting expectations were those for which the difference between the actual number of students and the expected number of students at or above expectations was not statistically significant. A 95% confidence interval was used; this criterion means that differences are only reported when there is a 5% or smaller probability that a difference of that size could occur by chance.

The results are reported in Table 4-6. The percentages in the table are based on 187 jurisdictions (including private schools).

Table 4-6
Grade 6 Social Studies
Percentage Distribution of Jurisdictions* Meeting Expectations

Reporting Category and Level of Standard	Below Expectations	Meeting Expectations	Above Expectations
Total Test			
Acceptable Standard**	29.4	64.2	6.4
Standard of Excellence***	6.4	73.8	19.8
Part A: Multiple Choice			
Acceptable Standard	20.2	69.0	10.2
Standard of Excellence	1.6	53.5	44.9
Part B: Written Response			
Acceptable Standard	61.0	39.0	0.0
Standard of Excellence	20.9	76.5	2.7

*Jurisdictions with fewer than five students are excluded, as the statistical significance of the frequencies compared to the expectations cannot be calculated.
**Acceptable Standard: 85% of students are expected to achieve at or above the acceptable standard.
***Standard of Excellence: 15% of students are expected to achieve at or above the standard of excellence.

AVERAGE SCORE

Another way to look at the achievement of students is by means of the average score.

The average score for the total Grade 6 Social Studies Achievement Test was 62.5%, with a standard deviation of 16.3.

Table 4-7 shows the percentage of students who obtained each total test score (relative frequency) and the percentage of students who scored at or below each total test score (cumulative frequency). Total test scores are expressed as percentages.

Table 4-7
Grade 6 Social Studies
Frequency Distribution of Total Test Scores

Total Score (%)	Relative Frequency (%)	Cumulative Frequency (%)	Total Score (%)	Relative Frequency (%)	Cumulative Frequency (%)
0	0.0	0.0	51	1.7	26.5
1	0.0	0.0	52	1.8	28.2
2	0.0	0.0	53	1.7	29.9
3	0.0	0.0	54	1.7	31.7
4	0.0	0.0	55	1.9	33.5
5	0.0	0.0	56	2.0	35.5
6	0.0	0.0	57	2.0	37.6
7	0.0	0.0	58	1.8	39.4
8	0.0	0.0	59	2.1	41.5
9	0.0	0.0	60	2.2	43.6
10	0.0	0.0	61	2.0	45.6
11	0.0	0.0	62	2.1	47.7
12	0.0	0.0	63	2.2	49.9
13	0.0	0.1	64	2.2	52.1
14	0.0	0.1	65	2.0	54.1
15	0.0	0.1	66	2.2	56.3
16	0.0	0.1	67	2.1	58.4
17	0.1	0.2	68	2.3	60.7
18	0.1	0.2	69	2.2	62.9
19	0.0	0.3	70	2.1	65.0
20	0.1	0.4	71	2.1	67.2
21	0.1	0.4	72	2.3	69.5
22	0.1	0.6	73	2.1	71.6
23	0.1	0.7	74	2.1	73.8
24	0.1	0.8	75	2.0	75.8
25	0.2	1.0	76	2.1	77.9
26	0.2	1.3	77	1.9	79.8
27	0.3	1.6	78	1.8	81.7
28	0.4	2.0	79	1.7	83.4
29	0.4	2.3	80	1.8	85.2
30	0.5	2.8	81	1.7	86.9
31	0.5	3.3	82	1.8	88.6
32	0.5	3.8	83	1.5	90.2
33	0.6	4.4	84	1.5	91.6
34	0.6	5.1	85	1.3	92.9
35	0.7	5.7	86	1.2	94.1
36	0.7	6.5	87	1.1	95.2
37	0.9	7.3	88	1.0	96.2
38	0.9	8.3	89	0.8	97.0
39	0.9	9.2	90	0.7	97.7
40	1.2	10.4	91	0.6	98.2
41	1.1	11.5	92	0.5	98.7
42	1.1	12.6	93	0.4	99.1
43	1.3	14.0	94	0.3	99.5
44	1.4	15.3	95	0.2	99.7
45	1.5	16.8	96	0.2	99.8
46	1.5	18.4	97	0.1	99.9
47	1.4	19.8	98	0.1	99.9
48	1.6	21.4	99	0.1	100.0
49	1.7	23.1	100	0.0	100.0
50	1.7	24.7			

RESULTS FOR PART A: MULTIPLE CHOICE

RESULTS IN RELATION TO STANDARDS

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63). For Part A: Multiple Choice, the standard established was such that in order to meet

- the acceptable standard, students had to achieve a raw score of 23 out of 50
- the standard of excellence, students had to achieve a raw score of 40 out of 50.

Based on these standards, the results revealed that

- 83.3% of students performed at or above the acceptable standard, and
- 24.3% of students performed at or above the standard of excellence.

The level of performance was slightly lower than was expected at the acceptable standard and was much higher than was expected at the standard of excellence.

AVERAGE SCORE

Provincial summary results for Part A: Multiple Choice were as follows:

- Provincial Average -- 32.2 marks out of a possible 50
- Standard Deviation -- 8.9

As outlined in the blueprint on page 22, the questions on Part A: Multiple Choice were grouped according to reporting categories.

Table 4-8, page 29, presents provincial averages for these reporting categories.

Provincial averages were computed and rounded to one decimal. Consequently, the sum of the averages for the reporting categories is not the same as the average for the total test.

Although levels of performance in the different reporting categories appeared to show some variation, caution is advised when comparing them. The sets of questions that made up each category were not selected to be equal in average level of difficulty; therefore, differences may have been due to variations in question difficulty rather than in student performance. Jurisdiction and school results can be usefully compared with the provincial averages to detect patterns of relative strength or weakness in achievement in each of the reporting categories.

Table 4-8 shows the averages and standard deviations for each of the reporting categories specified by the blueprint for Part A: Multiple Choice.

Table 4-8
Grade 6 Social Studies
Part A: Multiple Choice
Raw Score Results by Reporting Category

Reporting Category	Number of Questions	Raw Score Average	Standard Deviation
Topic A: How People in Earlier Times Met Their Needs	17	12.3	3.0
Topic B: How People in Eastern Societies Meet Their Needs Today	16	9.7	3.5
Topic C: Meeting Needs Through Local, Provincial, and Federal Governments	17	10.1	3.8
Knowledge and Comprehension All Topics	22	14.0	4.2
Topic A	7	5.5	1.4
Topic B	7	4.0	1.7
Topic C	8	4.5	2.2
Value Concepts and Valuing Skills (All Topics)	6	4.1	1.5
Inquiry Skills I (All Topics)	8	5.2	1.8
Inquiry Skills II (All Topics)	8	5.0	1.9
Inquiry Skills III (All Topics)	6	3.9	1.5

Table 4-9 shows the percentage of students who obtained each score on Part A: Multiple Choice (relative frequency) and the percentage of students who scored at or below each score (cumulative frequency).

Table 4-9
Grade 6 Social Studies
Part A: Multiple Choice
Frequency Distribution of Raw Scores

Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)	Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)
0	0.0	0.0	26	3.0	28.2
1	0.0	0.0	27	3.2	31.3
2	0.0	0.0	28	3.4	34.7
3	0.0	0.0	29	3.3	38.0
4	0.0	0.0	30	3.6	41.7
5	0.0	0.0	31	3.4	45.0
6	0.0	0.0	32	3.6	48.7
7	0.0	0.1	33	3.7	52.4
8	0.1	0.1	34	3.7	56.1
9	0.1	0.2	35	4.0	60.2
10	0.2	0.4	36	3.8	63.9
11	0.3	0.7	37	4.0	68.0
12	0.5	1.2	38	4.0	72.0
13	0.6	1.8	39	3.7	75.7
14	0.8	2.6	40	3.6	79.3
15	1.0	3.5	41	3.6	82.9
16	1.2	4.8	42	3.3	86.2
17	1.4	6.2	43	3.2	89.4
18	1.8	8.0	44	2.9	92.2
19	1.9	9.9	45	2.6	94.8
20	2.1	12.0	46	2.0	96.8
21	2.2	14.2	47	1.4	98.2
22	2.5	16.7	48	1.0	99.3
23	2.7	19.4	49	0.5	99.8
24	2.8	22.2	50	0.2	100.0
25	3.0	25.2			

PERCENTAGE OF STUDENTS CHOOSING EACH ALTERNATIVE

Table 4-10 presents the percentage of students who chose each alternative (A, B, C, and D) for each multiple-choice question on Part A: Multiple Choice. The correct response (key) for each question is also identified.

The results shown in Table 4-10 can best be used in conjunction with results presented in jurisdiction and school reports in order to interpret strengths and weaknesses of local programs.

Table 4-10
Grade 6 Social Studies
Part A: Multiple Choice
Results for Individual Multiple-Choice Questions*

		Distribution of Responses (%)						Distribution of Responses (%)			
Item	Key	A	B	C	D	Item	Key	A	B	C	D
1	D	2.9	12.6	8.1	76.3	26	A	51.2	16.4	5.6	26.7
2	B	4.0	64.6	8.6	2.7	27	C	18.6	9.3	65.0	7.0
3	A	76.5	8.3	9.2	5.9	28	A	60.9	15.7	10.1	13.1
4	A	81.2	6.5	4.3	7.8	29	D	16.4	15.8	13.7	53.8
5	D	4.4	8.1	11.1	76.2	30	B	6.2	76.3	9.2	7.9
6	B	3.1	79.9	8.0	8.9	31	D	10.1	16.6	12.4	60.7
7	C	12.2	11.4	70.3	5.9	32	D	11.9	14.8	13.7	59.3
8	B	2.7	87.6	5.8	3.8	33	A	56.6	12.8	6.9	23.2
9	A	84.0	2.0	10.8	3.1	34	B	8.5	72.4	10.5	8.2
10	D	7.0	5.4	1.9	85.7	35	D	4.3	8.8	18.8	67.6
11	A	68.7	10.3	9.4	11.4	36	C	7.6	12.6	65.9	13.2
12	C	18.6	17.0	53.2	11.1	37	A	31.7	12.0	11.8	43.8
13	B	22.7	63.3	6.0	7.8	38	B	11.0	64.2	20.0	3.9
14	C	6.7	16.4	68.7	8.0	39	C	8.2	8.3	61.4	21.0
15	B	6.4	56.1	22.7	14.6	40	A	74.2	8.4	10.4	5.8
16	C	15.7	10.8	64.9	8.5	41	D	5.0	9.7	25.0	58.8
17	D	5.2	7.9	16.7	70.0	42	D	7.2	10.0	8.4	72.8
18	B	14.2	69.2	4.6	11.9	43	D	28.5	9.6	7.0	53.0
19	A	79.8	9.6	5.9	4.6	44	B	22.5	43.0	23.3	9.1
20	C	11.4	8.5	39.9	40.1	45	A	43.4	13.9	24.0	16.6
21	A	40.3	13.8	15.7	30.0	46	D	22.4	11.5	5.0	58.7
22	C	15.7	11.4	63.8	9.1	47	B	8.6	63.5	10.2	15.2
23	B	15.5	69.7	9.6	5.1	48	B	20.5	58.5	6.7	11.6
24	A	63.3	15.2	13.8	7.6	49	C	13.7	12.1	50.3	21.2
25	C	22.2	4.9	64.9	7.9	50	D	12.2	5.9	4.2	75.0

*The sum of the percentages for each question may be less than 100% because the No Response category is not included. The No Response category does not exceed 2.7% for any one of these questions.

RESULTS FOR PART B: WRITTEN RESPONSE

RESULTS IN RELATION TO STANDARDS

The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63). For Part B: Written Response, the standard established was such that in order to meet

- the acceptable standard, students had to achieve a raw score of 15 out of 30
- the standard of excellence, students had to achieve a raw score of 24 out of 30.

Based on these standards, the results revealed that

- 72.9% of students performed at or above the acceptable standard, and
- 11.9% of students performed at or above the standard of excellence.

Both levels of performance were much lower than were expected. See Examiners' Remarks, page 34, for a comparison with 1985.

AVERAGE SCORE

The average raw score for Part B: Written Response was 17.5 marks out of a possible 30, with a standard deviation of 5.1.

The results for each written-response question are summarized in Table 4-11.

Table 4-11
Grade 6 Social Studies
Part B: Written Response
Average Scores Awarded

Question	Total Marks Possible	Average Score	Difficulty Level*
Short Answer			
1	2	1.24	0.62
2	2	1.15	0.57
3	2	0.87	0.44
4	2	0.73	0.36
5	3	2.25	0.75
6	4	3.08	0.77
Composition			
7	15	8.19	0.55

*The difficulty level is the average score divided by the total marks possible.

Table 4-12 presents the distribution of scores for the six short-answer questions.

Table 4-12
Grade 6 Social Studies
Part B: Written Response
Distribution of Scores for Short-answer Questions

Question	Percentage of Students Obtaining Each Mark				
	0	1	2	3	4
1	24.3	27.8	47.8	N/A	N/A
2	33.9	17.8	48.3	N/A	N/A
3	42.7	27.3	30.0	N/A	N/A
4	50.1	27.3	22.6	N/A	N/A
5	10.5	13.4	16.4	59.7	N/A
6	3.3	10.9	13.7	19.1	53.0

Table 4-13 presents the distribution of scores for the one composition question.

Table 4-13
Grade 6 Social Studies
Part B: Written Response
Distribution of Scores for Composition Question

Scale Points	Persuasiveness and Logic (%)	Language and Expression (%)
5 (Excellent)	4.1	4.9
4 (Good)	14.4	18.4
3 (Satisfactory)	35.8	46.2
2 (Limited)	35.4	23.8
1 (Poor)	8.8	4.6
0 (blank paper, off topic, insufficient response, illegible)	1.6	2.0

GRADE 6 SOCIAL STUDIES
EXAMINERS' REMARKS

With the exception of the percentage of students achieving the acceptable standard on Part B: Written Response, which remained approximately the same as in 1985, more students achieved the acceptable standard and the standard of excellence in 1989 than in 1985. (See Appendix C, page 63.) Thus, while only 72.9% achieved the acceptable standard on Part B: Written Response, and 11.9% achieved the standard of excellence, this compares favorably with the results achieved in 1985.

The Part A: Multiple Choice results were almost as high as were expected. At the provincial level, students appear to be mastering satisfactorily the knowledge, skills, and value objectives of the curriculum. The percentage of students achieving the standard of excellence on the multiple-choice part is particularly encouraging.

The Part B: Written Response assignment required students to follow the steps of the inquiry process to address the issue of whether "the government should continue to help unemployed people meet their basic needs." This issue was chosen for its curricular fit and topicality, and because students could legitimately take either side. As well, the topic permitted assessment of students' ability to deal with abstract concepts.

As in 1985, the majority of students appeared to experience difficulty in forming a generalization. When asked in question 4 "What general statement can be made about meeting needs when a person is unemployed?", many students responded with advice for the unemployed rather than with a generalization that drew a relationship between unemployment and meeting basic needs.

Most students (89%) were able to identify the value positions underlying a speaker's statement, and to identify which speakers held opposing positions. Fewer students (60%) were able to provide evidence from the speakers' statements to demonstrate that the speakers held these values. Some students misunderstood the task required of them, and instead of providing evidence of the speakers' values, offered their own opinion on the issue.

The "evidence" portion of questions 5 and 6 created difficulty for markers. Markers were instructed to distinguish between those papers that included some superfluous editorializing and those in which extraneous material was sufficiently intrusive that students failed to address the assigned task. Inter-rater reliability was lowest on these two items, which suggests that there was some variation in where individual markers drew this line. This seems to reflect a fundamental disagreement between those markers who felt that students who offered their own opinion were "going beyond the task required" but deserved marks for effort, and those who felt that these particular students had failed to address the assigned task and therefore should receive a zero.

Papers that received a zero for question 7, the composition question, were reviewed. Of the 470 tests, 375 were blank for that question and two were illegible. The remaining 93 were off topic.

Markers commented that some students lost marks because they confused the issue of unemployment with retirement. This may have been the result of students overgeneralizing from practice exercises with the 1985 achievement test, which featured the issue of how best to care for the aged. Other students lost marks because they assumed the unemployed were disabled.

SECTION 5
GRADE 9 SCIENCE

GENERAL DESCRIPTION

The Grade 9 Science Achievement Test consisted of 75 multiple-choice questions. The time allotted for writing the test was 90 minutes.

The information presented in this section is based on the results achieved by 27 137 students.

SUMMARY OF RESULTS

Results in Relation to Standards

Results show that 80.4% of students who wrote the test achieved the acceptable standard and 22.9% achieved the standard of excellence. These results were lower than were expected for the acceptable standard and were higher than were expected for the standard of excellence. The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Average Score

The average total test score was 66.8%, with a standard deviation of 17.7. The average total test raw score was 50.1 marks out of a possible 75, with a standard deviation of 13.3.

CONTENT OF THE TEST

The Grade 9 Science Achievement Test was designed to reflect the *Grade 9 Science Curriculum Specifications* (revised May 1986). However, the scope of the test was limited to curriculum objectives that could be efficiently measured on a paper and pencil test. As a result, questions on the test were drawn from the content of the two major components in the core program:

- Subject Matter
- Process Skills

The subject matter component consisted of questions associated with the major concepts of the study of physical science in Grade 9. The process skills component consisted of questions integrated with subject matter and questions independent of subject matter.

Questions on the Grade 9 Science Achievement Test measured student achievement at three cognitive levels:

- Knowledge -- recognize or recall ideas, terminology, facts, conventions, methods of inquiry, principles, generalizations, theories, and concepts

- Comprehension and Application -- demonstrate an understanding of the concepts and skills, and apply appropriate methods and ideas to a new situation

- Higher Mental Activities -- demonstrate an ability to analyse and synthesize data in an effort to make generalizations, and evaluate ideas, solutions, and information.

Table 5-1 presents the blueprint used to develop the Grade 9 Science Achievement Test.

Classification of each question by component, subtest, and cognitive level is indicated in the table.

Table 5-1
Grade 9 Science
Achievement Test Blueprint

Reporting Category	Cognitive Levels*			Number of Questions	Component		Number of Questions
	K	C/A	HMA		Process Skills	Subject Matter	
Matter Occupies Space	*1,4, 10*	*2,3,5,6,8, 11,12,13*	*7,9*	13	*2,3,7, 9,11*	*1,4,5,6,8, 10,12,13*	13
Kinetic Molecular Theory	*14, 18, 23, 29*	*15,16,17, 19,20,21, 24,25,26, 27,30,31, 32*	*22, 28, 33*	20	*21,22,24, 26,28,32, 33*	*14,15,16, 17,18,19, 20,23,25, 27,29,30, 31*	20
Heat and Temperature	*34, 35, 46, 48*	*36,37,38, 40,41,42, 43,45,47, 49,51*	*39, 44, 50*	18	*36,43,44, 45,50,51*	*34,35,37, 38,39,40, 41,42,46, 47,48,49*	18
Energy	*52*	*53,55,56, 57,58*	*54*	7	*57,58*	*52,53,54, 55,56*	7
Atoms and Molecules	*65*	*59,60,61, 62,63,66, 67*	*64*	9	*66,67*	*59,60,61, 62,63,64, 65*	9
Process Skills as Content	*68, 69*	*70,71,72, 73,75*	*74*	8	*68,69,70, 71,72,73, 74,75*		8
Total	15	49	11	75	30	45	75

*K - - Knowledge
C/A - - Comprehension and Application
HMA - - Higher Mental Activities

STUDENTS TESTED, ABSENT, AND EXEMPT

The Grade 9 Science Achievement Test was available both in English and in French translation. Table 5-2 presents the number of students who wrote the Grade 9 Science Achievement Test or its French translation (9e Année Test de Rendement Sciences).

Table 5-2
Grade 9 Science
Students Tested

Type of Participation	Number of Students
Participation Required (Students Receiving Instruction in English)	27 137*
Participation Optional (Students Receiving Instruction in French)	
Wrote in English	64**
Wrote French Translation	1 036***

*Of the total number of students required to write the test, 1 205 students were absent the day the test was written and 1 318 students were exempt from writing the test. (See Table 5-3.)

**Results achieved by these students are not included in the provincial data because participation in the Achievement Testing Program is optional for these students.

***Results achieved by these students will be presented in a special report.

Table 5-3 presents the number and percentage distribution of students who were required to write the Grade 9 Science Achievement Test and who were absent or exempt.

Table 5-3
Grade 9 Science
Students Included in Provincial Results, Absentees, and Exemptions

Category	Number of Students	Percentage of Students
Students Included in Provincial Results	27 137	91.5
Students Absent	1 205	4.1
Students Exempt:	1 318	4.4
Special Needs	716	2.4
Not Taught Subject This Term	297	1.0
ESL Classes	166	0.5
Language of Instruction Not English	19	0.1
Other (approved by Director)	120	0.4

RESULTS FOR THE TOTAL TEST

RESULTS IN RELATION TO STANDARDS

Table 5-4 shows the percentage of students achieving the acceptable standard and the standard of excellence. These levels of performance were lower than were expected for the acceptable standard and were higher than were expected for the standard of excellence. The acceptable standard and the standard of excellence were established by a standard-setting procedure (see Appendix C, page 63).

Table 5-4
Grade 9 Science
Percentage of Students Achieving Standards

Level of Standard	Raw Score Representing Standard*	Percentage of Students Achieving At or Above Standard	
		Expected	Actual
Acceptable Standard	38	85.0	80.4
Standard of Excellence	62	15.0	22.9

*The maximum possible raw score was 75.

The numbers of students achieving the acceptable standard and the standard of excellence for each jurisdiction were analysed to determine whether jurisdictions were below expectations, meeting expectations, or above expectations. Jurisdictions classified as meeting expectations were those for which the difference between the actual number of students and the expected number of students at or above expectations was not statistically significant. A 95% confidence interval was used; this criterion means that differences are only reported when there is a 5% or smaller probability that a difference of that size could occur by chance.

The results are reported in Table 5-5. The percentages in the table are based on 173 jurisdictions (including private schools).

Table 5-5
Grade 9 Science
Percentage Distribution of Jurisdictions* Meeting Expectations

Level of Standard	Below Expectations	Meeting Expectations	Above Expectations
Total Test			
Acceptable Standard**	30.6	64.2	5.2
Standard of Excellence***	3.5	57.8	38.7

*Jurisdictions with fewer than five students are excluded, as the statistical significance of the frequencies compared to the expectations cannot be calculated.
**Acceptable Standard: 85% of students are expected to achieve at or above the acceptable standard.
***Standard of Excellence: 15% of students are expected to achieve at or above the standard of excellence.

AVERAGE SCORE

Another way to look at the achievement of students is by means of the average score.

The average score for the total Grade 9 Science Achievement Test was 66.8%, with a standard deviation of 17.7.

REPORTING CATEGORIES

Table 5-6 shows the total marks possible and the provincial raw score results for the reporting categories of the Grade 9 Science Achievement Test.

It is important to stress that the averages on the various reporting categories cannot be directly compared with one another. Rather, the results shown in Table 5-6 can best be used in conjunction with parallel tables in the jurisdiction, school, and classroom reports. Variations in patterns of students' responses to questions can help to indicate strengths and weaknesses in local educational programs.

Table 5-6
Grade 9 Science
Raw Score Results by Reporting Categories

Reporting Category	Total Marks Possible	Raw Score Average	Standard Deviation
Total Test	75	50.1	13.3
Major Components			
Subject Matter	45	29.5	8.2
Process Skills	30	20.6	5.7
Subtests			
Matter Occupies Space	13	8.3	2.7
Kinetic Molecular Theory	20	13.8	3.9
Heat and Temperature	18	12.0	3.8
Energy	7	4.9	1.7
Atoms and Molecules	9	5.7	2.2
Process Skills as Content	8	5.4	1.8
Cognitive Levels			
Knowledge	15	10.3	2.9
Comprehension/Application	49	33.0	9.0
Higher Mental Activities	11	6.8	2.4

FREQUENCY DISTRIBUTION OF RAW SCORES

Table 5-7 presents the percentage of students who obtained each score on the Grade 9 Science Achievement Test (relative frequency) and the percentage of students who scored at or below each score (cumulative frequency).

Table 5-7
Grade 9 Science
Frequency Distribution of Raw Scores

Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)	Total Raw Score	Relative Frequency (%)	Cumulative Frequency (%)
0	0.0	0.0	38	1.7	21.3
1	0.0	0.0	39	1.9	23.2
2	0.0	0.0	40	1.8	25.1
3	0.0	0.0	41	1.8	26.8
4	0.0	0.0	42	2.0	28.9
5	0.0	0.0	43	2.2	31.1
6	0.0	0.0	44	2.1	33.1
7	0.0	0.0	45	2.3	35.5
8	0.0	0.0	46	2.3	37.7
9	0.0	0.0	47	2.3	40.0
10	0.0	0.0	48	2.6	42.6
11	0.0	0.0	49	2.5	45.1
12	0.0	0.0	50	2.6	47.7
13	0.0	0.1	51	2.5	50.2
14	0.0	.01	52	2.7	52.9
15	0.1	0.2	53	2.6	55.5
16	0.1	0.3	54	2.7	58.2
17	0.2	0.5	55	2.6	60.8
18	0.2	0.7	56	2.8	63.7
19	0.3	1.0	57	2.6	66.3
20	0.4	1.4	58	2.8	69.1
21	0.5	1.9	59	2.6	71.7
22	0.5	2.4	60	2.8	74.5
23	0.5	2.9	61	2.6	77.1
24	0.7	3.6	62	2.5	79.6
25	0.8	4.4	63	2.3	81.9
26	0.8	5.2	64	2.5	84.4
27	0.9	6.2	65	2.4	86.7
28	1.0	7.2	66	2.4	89.1
29	1.0	8.2	67	2.1	91.2
30	1.2	9.4	68	1.9	93.1
31	1.3	10.8	69	1.8	94.9
32	1.3	12.1	70	1.7	96.5
33	1.3	13.4	71	1.3	97.8
34	1.4	14.8	72	1.0	98.8
35	1.5	16.3	73	0.7	99.5
36	1.6	17.9	74	0.4	99.9
37	1.7	19.6	75	0.1	100.0

PERCENTAGE OF STUDENTS CHOOSING EACH ALTERNATIVE

Table 5-8 shows the percentage of students who chose each alternative (A, B, C, and D) for each multiple-choice question. The correct response (key) for each question is also identified.

The results shown in Table 5-8 can best be used in conjunction with the parallel tables in the jurisdiction and school reports. Variations in patterns of students' responses to questions can help to indicate strengths and weaknesses in local educational programs.

Table 5-8
Grade 9 Science
Results for Individual Multiple-Choice Questions*

Item	Key	Distribution of Responses (%) A	B	C	D	Item	Key	Distribution of Responses (%) A	B	C	D
1	A	73.6	13.2	1.6	11.5	39	B	13.1	51.9	26.3	8.5
2	C	1.8	37.3	52.1	8.8	40	C	7.2	31.2	55.5	6.0
3	A	80.1	9.9	5.7	4.3	41	D	9.9	16.7	13.3	60.1
4	D	3.9	5.5	6.1	84.5	42	C	9.5	9.2	67.1	14.1
5	C	29.2	7.7	54.7	8.3	43	A	67.9	5.9	11.1	15.0
6	C	15.5	13.2	63.9	7.2	44	B	5.8	72.3	14.5	7.3
7	C	15.6	12.3	51.9	20.0	45	A	59.7	24.6	7.0	8.6
8	B	21.4	52.0	9.3	17.2	46	B	22.3	53.4	16.0	8.1
9	A	74.6	4.6	16.4	4.4	47	C	12.4	10.7	67.2	9.5
10	A	68.3	17.7	9.8	4.0	48	D	6.8	16.8	10.7	65.6
11	D	10.1	21.1	9.0	59.5	49	A	77.3	5.5	6.1	11.1
12	C	13.8	15.6	67.7	2.9	50	A	59.3	25.5	11.9	3.1
13	A	51.4	12.2	16.4	19.7	51	D	3.6	8.7	14.7	72.8
14	C	7.2	5.4	82.1	5.3	52	D	12.9	6.5	9.5	71.0
15	D	4.7	6.4	3.9	85.0	53	C	12.0	5.9	70.2	11.7
16	B	12.2	76.6	5.2	5.8	54	A	50.1	16.6	10.8	22.4
17	C	13.5	15.1	55.6	15.7	55	C	9.6	9.9	74.4	6.0
18	B	19.1	48.0	8.5	24.3	56	B	7.7	74.9	6.8	10.4
19	D	5.4	15.3	3.8	75.5	57	D	5.3	5.1	16.6	72.8
20	C	5.6	9.1	66.5	18.7	58	D	8.2	10.6	6.7	74.3
21	D	4.1	14.0	6.5	75.4	59	B	8.6	68.4	8.7	14.2
22	C	2.6	22.4	72.2	2.8	60	B	17.2	56.0	11.5	15.2
23	A	73.4	8.7	7.8	10.0	61	B	9.2	33.1	33.3	24.2
24	A	73.3	7.7	8.1	10.9	62	D	4.7	11.1	16.9	67.1
25	C	4.9	15.5	75.2	4.4	63	D	11.0	15.5	10.8	62.5
26	B	11.1	75.3	4.4	9.1	64	B	11.2	55.9	8.9	23.8
27	B	8.9	79.2	5.6	6.3	65	C	12.4	8.6	69.4	9.4
28	B	14.4	49.4	21.2	14.8	66	B	5.5	81.6	9.5	3.2
29	B	26.9	48.2	15.1	9.7	67	D	7.5	13.0	5.7	73.6
30	A	64.6	18.7	9.6	6.9	68	B	9.4	77.8	7.2	5.4
31	C	7.9	10.1	69.0	12.9	69	D	15.1	13.0	10.9	60.8
32	D	17.3	3.6	8.7	70.3	70	B	9.5	71.9	5.7	12.6
33	A	63.8	2.1	7.9	26.1	71	D	19.2	6.4	10.4	63.7
34	A	80.2	11.7	3.4	4.7	72	B	13.4	46.1	17.9	22.3
35	A	75.6	4.1	9.8	10.4	73	D	3.1	12.3	3.8	80.6
36	B	6.3	80.4	4.4	8.8	74	B	6.9	79.9	5.9	7.0
37	C	5.4	10.3	73.1	11.2	75	A	63.8	15.0	13.4	7.4
38	C	18.5	18.3	55.8	7.1						

*The sum of the percentages for each question may be less than 100% because the No Response category is not included. The No Response category does not exceed 0.4 % for any one of these questions.

GRADE 9 SCIENCE
EXAMINERS' REMARKS

Grade 9 Science teachers who were chosen as markers and standard-setters for the 1989 Grade 9 Science Achievement Test felt that, generally, the test reflected the essence of the Grade 9 Science Program. In addition, teachers agreed that the test presented many questions that required students to think and to apply both science process and inquiry skills to analyse real-life situations. Standard-setting results in Table 7-6, page 55, indicate that in 1989, 80.4% of students met the acceptable standard. Results indicate that slightly more students, 81.3%, met the acceptable standard in 1985. In essence, the number of students achieving the acceptable standard has not changed appreciably from 1985 to 1989.

In 1989, 22.9% of students met or exceeded the standard of excellence. However, comparable standards set in 1989 for the 1985 test indicate that 19.5% of students met or exceeded the standard of excellence. This suggests that a slightly higher number of students achieved at or above the standard of excellence in 1989 than in 1985.

Teacher-markers felt that students were generally comfortable with the reading level of the questions. Students' strengths and weaknesses in both subject matter and process skill questions are discussed below.

Question 2 required students to identify the correct measurement for the line below a centimeter ruler. Many students (37.3%) chose 10.6 cm instead of 9.6 cm as the correct answer. This suggests that students failed to recognize that the line started at the 1.0 cm mark. As a result, students may have read the length directly above the line on the ruler without subtracting 1.0 cm from 10.6 cm.

Question 7 required students to follow a four-step procedure to calculate the volume of a glass rod. A difficulty level of 51.9% indicates that some students experienced problems in sorting out relevant and irrelevant data.

Question 36 required students to identify the revision that would most likely improve an experiment. Most (80.4%) had little difficulty choosing the correct revision. This result seems to indicate that most Grade 9 Science students have little trouble analysing data and revising experiments of this type.

Question 45 (difficulty level of 59.7%) required students not only to interpret from a graph but also to identify the process that might take place if heat were removed from Material X at 160°C. The results for questions 3 (difficulty level of 80.1%) and 27 (difficulty level of 79.2%), which also concern graphs, suggest that interpreting graphs is a skill that most Grade 9 Science students possess. The difficulty they had with question 45, therefore, may be the inability of many students to integrate their knowledge and skills and transfer them to a new situation.

When comparing the results of the 1989 and 1985 Science achievement tests, it seems that, although students have shown some improvement in their ability to recall and/or recognize facts, ideas, concepts, and generalizations, overall student achievement in 1989 appears to be unchanged compared to 1985.

SECTION 6
STUDENT ACHIEVEMENT BY GENDER

The figures and tables in this section of the report provide a breakdown by gender of the results achieved by students who wrote the June 1989 achievement tests.

This information is a descriptive report of male and female student achievement in Grade 3 English Language Arts, Grade 6 Social Studies, and Grade 9 Science. It is a report of *what* happened in the June 1989 Achievement Testing program.

The interpretation of this information -- hypothesizing *why* results are as they are -- requires thoughtful consideration of the numerous variables that contribute to achievement. (See Section 2, page 3, and Appendix A, page 57, to interpret these results.)

Figure 6-1 shows the number of students by gender who wrote achievement tests in June 1989.

Figure 6-1

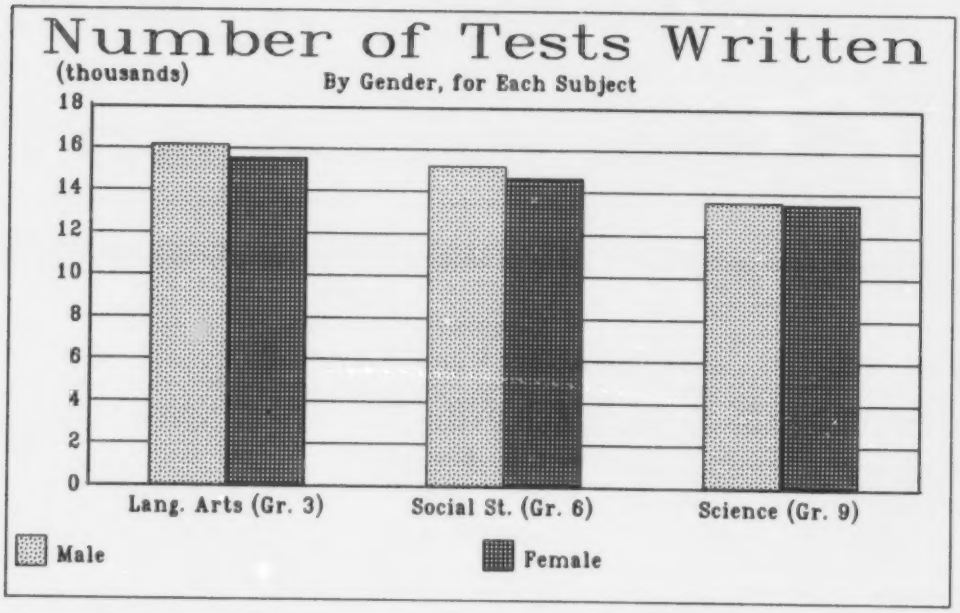


Figure 6-2 presents the percentage of students achieving the acceptable standard by gender.

Figure 6-2

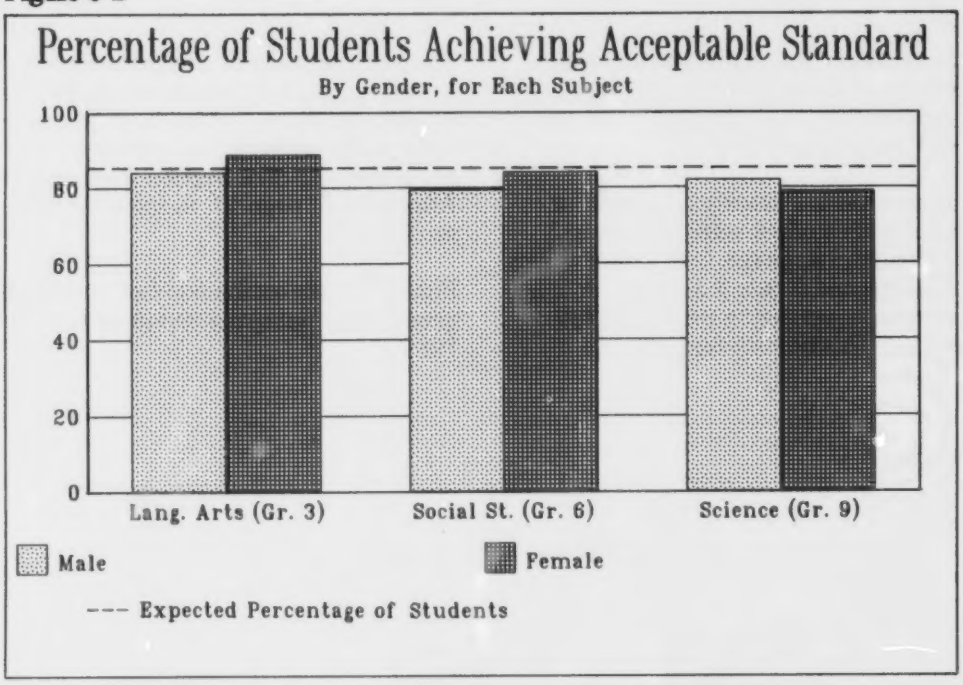


Figure 6-3 presents the percentage of students achieving the standard of excellence by gender.

Figure 6-3

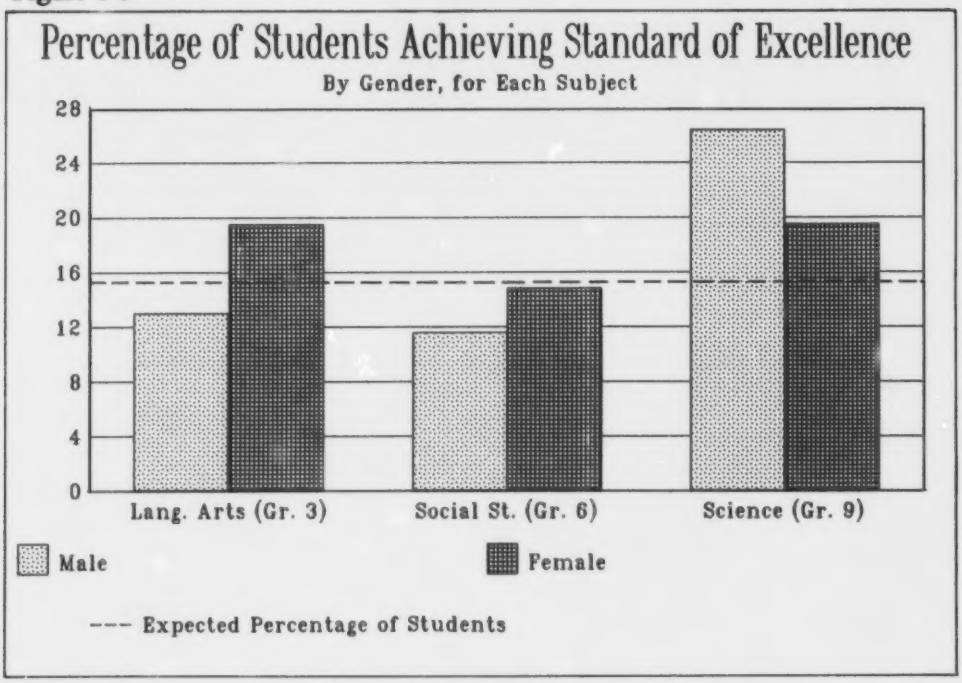


Table 6-1 presents the number and percentage of students by gender who wrote the June 1989 achievement tests.

Table 6-1
Number and Percentage of Students Writing Achievement Tests by Gender

Subject	Number of Students	Percentage of Students
Grade 3 English Language Arts ***N*=31 649***		
Male	16 152	51.0
Female	15 497	49.0
Grade 6 Social Studies ***N*=29 788***		
Male	15 178	51.0
Female	14 610	49.0
Grade 9 Science ***N*=27 008***		
Male	13 557	50.2
Female	13 451	49.8

*Gender was not reported for all students; therefore, totals differ slightly from those given elsewhere in this report.

Table 6-2 presents average scores and standard deviations by gender.

Table 6-2
Grades 3, 6, and 9 Achievement Tests
Average and Standard Deviation by Gender

Test	Total Score (%)		Multiple Choice (Raw Score)		Written Response (Raw Score)	
	Male	Female	Male	Female	Male	Female
Grade 3 English Language Arts			(Max Possible=40)		(Max Possible=25)	
Average	67.1	70.8	28.5	29.7	15.7	16.8
Standard Deviation	15.9	15.5	8.3	7.8	4.0	4.0
Grade 6 Social Studies			(Max Possible=50)		(Max Possible=30)	
Average	61.3	63.8	31.7	32.5	16.7	18.4
Standard Deviation	16.4	16.1	9.1	8.9	5.1	5.0
Grade 9 Science			Grade 9 Science consists of multiple-choice questions only.			
Average	68.2	65.4				
Standard Deviation	18.1	17.4				

SECTION 7
STUDENT ACHIEVEMENT OVER TIME

INTRODUCTION

PURPOSE

An important goal of Alberta Education is to measure and report changes in student achievement. Comparing student performance on the 1985 and the 1989 achievement tests is one way of meeting this goal. A direct comparison of the average scores on the 1985 and the 1989 tests cannot be made; although the achievement tests were parallel in form and content, their levels of difficulty may not have been equal. A study was therefore undertaken to permit comparisons of student performance in 1985 and in 1989.

STUDY DESIGN

The study consisted of two parts for the Grade 3 English Language Arts and Grade 6 Social Studies tests: test equating and standard setting. The Grade 9 Science study consisted of three parts: test equating, standard setting, and comparing equivalent items.

To equate the 1985 and the 1989 test scores, samples of grades 3, 6, and 9 students who completed the 1989 achievement tests also wrote the 1985 Grade 3 English Language Arts, Grade 6 Social Studies, and Grade 9 Science achievement tests. The students selected for the study were from schools chosen to be representative of rural, urban, large, and small schools throughout the province.

The 1985 and the 1989 multiple-choice scores from these students were used to establish what the 1985 equivalent scores would be on the 1989 test. (For an explanation of the multiple-choice test-equating process, refer to Appendix D, page 64.) Based on this equating process, the scores for the 1989 population were converted to estimated equivalent 1985 scores, and the 1985 and the 1989 populations were then compared.

An additional procedure was required to equate the written-response sections of the 1985 and the 1989 English Language Arts and Social Studies achievement tests (see Appendix E, page 65.)

The standard-setting part of the study involved groups of experienced teachers of grades 3, 6, and 9. Their task was to establish the scores that represent the acceptable standard and the standard of excellence on each of the three achievement tests in 1985 and in 1989. The percentage of students meeting each standard for the respective tests was then compared.

In addition to test equating and standard setting, a third procedure was used to compare student achievement in Grade 9 Science. This procedure involved comparing achievement on items in 1989 that were equivalent to those of 1985. Seven items from the 1985 test were paired with seven equivalent items from the 1989 test. The average score obtained by the 1985 and the 1989 populations of students on their respective sets of equivalent items was compared.

LIMITATIONS OF THE STUDY

Many factors other than changes in levels of achievement could have contributed to differences in student performance on the 1985 and the 1989 achievement tests. The Achievement Testing Program has been in place since 1982, and the 1985 tests were the first for those particular combinations of grade and subject. All those concerned have had more experience and had opportunity to adjust to the program since 1985. Content, understanding, and use of the bulletin provided by the Student Evaluation and Records Branch may have resulted in improved preparation for testing. Teachers in 1989 would have had an opportunity to use the 1985 tests for practice and to familiarize students with test-writing techniques. This opportunity did not exist in 1985. The effects of these extraneous variables could not be controlled; caution is therefore required when drawing conclusions about the results of the study.

GRADE 3 ENGLISH LANGUAGE ARTS

In 1989, the Grade 3 English Language Arts Achievement Test was administered to 34 090 students enrolled in the Grade 3 English Language Arts program.

The 1985 and the 1989 Grade 3 English Language Arts achievement tests consisted of two parts, Part A: Writing (the 1985 booklet was entitled Part A: Composition) and Part B: Reading. On the 1985 test, Part B: Reading contained 36 multiple-choice questions; on the 1989 test, there were 40 multiple-choice questions.

The average total test score in 1985 was 66.9%; the average score in 1989 was 68.9%.

The special comparison study involved administering the 1985 test to a sample of students who were also writing the 1989 test and using a single group of judges to determine the standards for both tests.

COMPARISON THROUGH RE-ADMINISTERING THE 1985 TEST

In June 1989, 215 Grade 3 students from eight schools in eight jurisdictions were selected for test-equating purposes. (For an explanation of the test-equating process, refer to Appendix D, page 64, and Appendix E, page 65.) The schools were chosen to be representative of rural, urban, large, and small schools throughout the province. Based on the results of the test-equating process, the total scores for the 1989 population were converted to estimated equivalent 1985 scores, and the 1985 and the 1989 populations were then compared. Table 7-1 shows the results of the comparison.

Table 7-1
Grade 3 English Language Arts
Comparison of 1985 and 1989 Test Results

Category	1985 Test	1989 Test
Number of Marks	100	100
Number of Students Writing	32 110	34 090
Average Score after Equating	66.9%	68.5%
Standard Deviation after Equating	14.4	14.8

The average 1989 score was higher than the average 1985 score by 1.6 marks out of a possible 100. A two-tailed t-test showed that this difference was statistically significant beyond the 0.001 level of probability: therefore, the difference is real and large enough to be of practical significance.

STANDARD SETTING

Twenty experienced Grade 3 English Language Arts teachers from schools throughout the province were selected to participate in the standard-setting section of the study. These teachers were among those marking the written-response part of the 1989 test. (See Appendix C, page 63, for an explanation of the standard-setting procedure.)

The teachers reviewed the multiple-choice parts of both 1985 and 1989 tests. They set scores that would represent the standards for each test: the acceptable standard and the standard of excellence. Table 7-2 shows the results of the standard-setting process.

Table 7-2
Grade 3 English Language Arts
Comparison of Percentage of Students Achieving Standards
1985 and 1989

Test	Year	Maximum Score	Acceptable Standard		Standard of Excellence	
			Score Representing Standard	Percentage of Students Achieving At or Above Standard	Score Representing Standard	Percentage of Students Achieving At or Above Standard
Total	1985	100	51	87.1	82	15.0
	1989	100	51	86.2	85	16.1
Part A: Writing	1985	25	13	79.5	20	17.7
	1989	25	13	84.5	20	20.6
Part B: Reading	1985	36	18	89.2	30	28.8
	1989	40	21	82.7	36	25.4

The results revealed that nearly the same number of students met the standards in 1985 and in 1989. Since a change of one in the raw score representing the standard will result in a change of approximately four per cent in the percentage of students meeting the standards, the differences shown are not important.

The standard-setting results showed that, in the judgment of the standard-setters, students achieved the acceptable standard and the standard of excellence as frequently in both test administrations.

CONCLUSION

The results of the test-equating analysis (shown in Table 7-1) indicate that achievement in those aspects of language arts that were tested was higher in 1989 than in 1985, but similar percentages of students achieved acceptable or excellent standards (shown in Table 7-2). Because the test-equating process yields more precise results than the standard-setting process, it was concluded that achievement in language arts was somewhat higher in 1989 than in 1985.

GRADE 6 SOCIAL STUDIES

In 1989, the Grade 6 Social Studies Achievement Test was administered to 29 995 students enrolled in the Grade 6 Social Studies program.

The 1985 and the 1989 Grade 6 Social Studies achievement tests each included 50 multiple-choice questions. The 1985 test had five written-response questions worth 30 marks, and the 1989 test had seven written-response questions worth 30 marks.

The average score in 1985 was 59.1%; the average score in 1989 was 62.5%.

The special comparison study involved administering the 1985 test to a sample of students who were also writing the 1989 test and using a single group of judges to determine the standards for both tests.

COMPARISON THROUGH RE-ADMINISTERING THE 1985 TEST

Ten schools in 10 jurisdictions were selected for the study. These schools were chosen to be representative of rural, urban, large, and small schools throughout the province. In June 1989, 244 Grade 6 students who were writing the 1989 test also wrote the 1985 test. Based on the results of the test-equating process, all total scores for the students who wrote tests in 1989 were converted to estimated equivalent 1985 scores, and the 1985 Grade 6 results and the 1989 Grade 6 results were then compared. (For an explanation of the test-equating process, refer to Appendix D, page 64, and Appendix E, page 65.) Table 7-3 shows the results of the comparison.

Table 7-3
Grade 6 Social Studies
Comparison of 1985 and 1989 Test Results

Category	1985 Test	1989 Test
Number of Marks	100	100
Number of Students Writing	29 551	29 995
Average Score after Equating	59.1%	59.3%
Standard Deviation after Equating	15.2	16.6

The average 1989 score was higher than the average 1985 score by 0.2 marks out of a possible 100. A two-tailed t-test showed that this difference was statistically significant beyond the 0.05 level of probability. However, the reliability of the estimated scores is limited. Considering this limitation, the difference does not appear to reflect a real difference in achievement scores.

STANDARD SETTING

Twenty-one experienced Grade 6 Social Studies teachers from schools throughout the province were selected to participate in the standard-setting section of the study. (See Appendix C, page 63, for an explanation of the standard-setting procedure.)

The teachers reviewed both 1985 and 1989 tests. They set scores that would represent the standards for each test: the acceptable standard and the standard of excellence. Table 7-4 shows the results of the standard-setting process.

Table 7-4
Grade 6 Social Studies
Comparison of Percentage of Students Achieving Standards
1985 and 1989

Test	Year	Maximum Score	Acceptable Standard		Standard of Excellence	
			Score Representing Standard	Percentage of Students Achieving At or Above Standard	Score Representing Standard	Percentage of Students Achieving At or Above Standard
Total	1985	100	47	78.4	80	8.9
	1989	100	47	81.4	80	16.6
Part A: Multiple Choice	1985	50	23	78.8	40	14.3
	1989	50	23	83.3	40	24.6
Part B: Written Response	1985	30	15	73.2	24	8.1
	1989	30	15	72.9	24	11.9

The results revealed that more students met the standards in 1989 than in 1985. Since a change of one in the raw score representing the standard will result in a change of approximately four per cent in the percentage of students meeting the standard, the differences shown are not important.

The standard-setting results showed that, in the judgment of the standard-setters, students achieved the acceptable standard and the standard of excellence as frequently in both test administrations.

CONCLUSION

The results of the test-equating analysis (shown in Table 7-3) indicate that achievement in those aspects of social studies that were tested was about the same in 1985 and in 1989, and similar percentages of students achieved the acceptable standards (shown in Table 7-4). Higher percentages of students met the standard of excellence in 1989 than in 1985. Because the test-equating process yields more precise results than the standard-setting process, it was concluded that achievement in social studies was essentially the same in 1989 as in 1985.

GRADE 9 SCIENCE

In 1989, the Grade 9 Science Achievement Test was administered to 27 201 students enrolled in the Grade 9 Science program.

The achievement tests in both 1985 and 1989 consisted of 75 multiple-choice questions.

The average score in 1985 was 66.2%; the average score in 1989 was 66.8%.

The special comparison study involved administering the 1985 test to a sample of 249 students who were also writing the 1989 test, using a single group of judges to determine the standards for both tests, and comparing equivalent questions.

COMPARISON THROUGH RE-ADMINISTERING THE 1985 TEST

Eight schools in eight jurisdictions were selected to participate in the study. These schools were chosen to be representative of rural, urban, large, and small schools throughout the province. In June 1989, 249 Grade 9 students who were writing the 1989 test also wrote the 1985 test. Based on the results of the test-equating process, all multiple-choice scores for the 1989 population were converted to estimated equivalent 1985 scores, and the 1985 and the 1989 populations were then compared. (For an explanation of the test-equating process, refer to Appendix D, page 64, and Appendix E, page 65, of this report.) Table 7-5 shows the results of the comparison.

Table 7-5
Grade 9 Science
Comparison of 1985 and 1989 Test Results

Category	1985 Test	1989 Test
Number of Questions	75	75
Number of Students Writing	29 639	27 201
Average Score after Equating	66.2%	65.5%
Standard Deviation after Equating	16.7	17.3

The average 1985 score was 0.7% higher than the average 1989 score. A two-tailed t-test showed that this difference was statistically significant beyond the 0.001 level of probability. However, this difference may not be educationally significant.

STANDARD SETTING

Twenty-one experienced Grade 9 Science teachers from schools throughout the province were selected to participate in the standard-setting section of the study. (See Appendix C, page 63, for an explanation of the standard-setting procedure.)

These teachers reviewed both 1985 and 1989 tests. They set scores that would represent the standards for each test: the acceptable standard and the standard of excellence. Table 7-6 shows the results of the standard-setting process.

Table 7-6
Grade 9 Science
Comparison of Percentage of Students Achieving Standards
1985 and 1989

Year	Maximum Score	Acceptable Standard		Standard of Excellence	
		Raw Score Representing Standard	Percentage of Students Achieving At or Above Standard	Raw Score Representing Standard	Percentage of Students Achieving At or Above Standard
1985	75	38	81.3	62	19.5
1989	75	38	80.4	62	22.9

The percentage of students meeting the acceptable standard in 1989 was virtually the same as in 1985.

However, the percentage of students meeting the standard of excellence was higher in 1989 than in 1985.

EQUIVALENT QUESTIONS

Seven questions from the 1989 achievement test were judged to be equivalent to seven questions on the 1985 achievement test by the Grade 9 Science Test Review Committee and by the 21 teachers who participated in the standard-setting process.

Questions were judged to be equivalent if they had all of the following characteristics:

1. They tested the same specific curriculum objectives.
2. They required the same level of reading skill.
3. The wording was similar in terms of complexity and the quantity of information presented.
4. They required the same number of steps or thought processes.
5. All three distractors were similar in terms of the process required to arrive at those answers.

The average score on the seven questions was 65.8% in 1985 and 67.2% in 1989.

CONCLUSION

The results obtained from equating the 1985 and the 1989 test results indicated that overall achievement in 1989 was virtually unchanged compared to 1985; however, standard-setting results indicated that a higher percentage of students achieved the standard of excellence in 1989 than in 1985. In addition, student achievement on the seven equivalent questions was slightly higher in 1989 than in 1985.

It can be concluded that student achievement in Grade 9 Science in 1989 was the same as in 1985.

APPENDIX A
USING ACHIEVEMENT TEST RESULTS

A SYSTEMATIC APPROACH FOR THE EFFECTIVE USE OF ACHIEVEMENT TEST RESULTS

Achievement test results can be used constructively as one means of improving the quality of education. A systematic use of achievement test results would include the following steps:

1. Comparing test results for a jurisdiction or school to the provincial results. Be sure that your comparisons include the

   - total test score,
   - total and subtest scores for multiple-choice questions,
   - total and subtest scores for written-response assignments (when appropriate),
   - individual multiple-choice question results, and
   - individual written-response question results (when appropriate).

2. Noting any patterns, anomalies, and/or interrelationships in the results.

3. Hypothesizing relationships between your observations and any of the factors listed in Section 2 of this report that may have had an effect on achievement or achievement test results.

4. Considering and implementing a plan that will help to improve the quality of education for students.

AN ADMINISTRATIVE MODEL FOR THE EFFECTIVE USE OF ACHIEVEMENT TEST RESULTS

The following model may be useful for those who wish to develop a constructive system for interpreting achievement test results. This model is based on work done by Medicine Hat School District #76.

BASIC PRINCIPLES

1. It is desirable and feasible for teachers and school administrators to make use of achievement test results in analysing the performance of their own students.

2. It is more constructive for schools to develop their own analyses, interpretations, and action plans than to have these imposed externally.

3. The impact of factors such as those listed in Section 2 should be analysed and discussed when reviewing achievement test results.

4. Subtest or reporting category results are usually more informative than total test scores.

5. Generalizations should be stated with caution and should be supported by evidence that is independent of achievement test results.

6. It is neither desirable nor productive to compare the results of different schools.

7. Achievement tests measure many of the objectives specified by the curriculum. However, skills and concepts that are not measured by the achievement tests are also to be taught and evaluated at the local level.

8. Staff discussions as well as written reports are useful means of ensuring that results are appropriately interpreted and used.

SUGGESTED CONTENT FOR INTERPRETATION OF INDIVIDUAL SCHOOL RESULTS

1. Subject, grade level, and date of achievement test administration
2. Number of students who wrote the achievement test
3. Profiles of students or groups who wrote the achievement test, which include

   •previous performances
   •number of students repeating the grade, etc.
4. Program emphases, such as hours of instruction, skill and content emphases
5. Instructional practices, such as methodology, resources, and the relationship between the program offered and the provincial curriculum
6. Program objectives not measured by the achievement test
7. School results compared to provincial results on subtests
8. Current school results compared to those of previous administration
9. Discussion of item results, identification of common student errors, and suggestions of ways for reducing the misunderstanding that leads to these errors
10. Recommendations for the following year or semester
11. Summary and concluding comments

SUGGESTED REPORTING STRUCTURE

1. Teachers and/or the principal analyse the results and prepare a written report about each administration of an achievement test.
2. The principal reviews and signs the report.
3. The report is shared with central office supervisory personnel.
4. The appropriate central office supervisory personnel prepare a written response to the report, with copies of the response going to the teachers and the principal.
5. If possible, all involved staff meet to discuss the report and the response.
6. Reports are used to improve the program and maximize future opportunities for student success.
7. When large differences exist between expected and actual achievement test results over time, consideration should be given to conducting a formal program evaluation.

APPENDIX B
DEVELOPING ACHIEVEMENT TESTS

The Student Evaluation and Records Branch develops achievement tests that measure student achievement at the grades 3, 6, and 9 levels. Provincewide testing in Language Arts, Mathematics, Science, and Social Studies follows a four-year cycle for each grade level and subject. Many individuals and groups are involved in the development of each test: practising classroom teachers, school and central office administrators, and representatives of postsecondary institutions, the Curriculum Design Branch, the Language Services Branch, Regional Offices, and the Student Evaluation and Records Branch. Student Evaluation and Records Branch staff ensure the development of valid and reliable tests.

Following is a summary of the phases of the test development process.

1. Planning
2. Approving Blueprints
3. Developing Test Questions
4. Constructing and Administering Field Tests
5. Analysing and Revising
6. Constructing Final Field Tests
7. Approving Final Field Tests
8. Administering Final Field Tests
9. Constructing the Final Test
10. Preparing and Administering the Final Test
11. Marking
12. Analysing and Reporting the Results

Under normal circumstances it takes three years to complete all phases of the process.

1. PLANNING

Test developers ensure that the design of each achievement test reflects the goals and objectives of the Program of Studies and the curriculum specifications for each subject. Planning takes into consideration those parts of the program that are testable in a paper and pencil format, within a given time frame. Teachers and consultants from across the province assist in preparing the design of each test.

Test developers prepare an interim test blueprint (an overall plan used to guide the development of a test). Questions that must be addressed at this point are:

- What knowledge and skills should students be expected to possess?
- What types of questions will constitute the test (multiple choice, short answer, or extended written response)?
- What weighting will each part of the test be given?
- How long and how demanding should the test be?
- How should the results of the test be organized for reporting purposes?

In order to ensure that each test will produce meaningful and reliable results, test developers incorporate statistical as well as curricular standards in the test design. Statistical standards include projected test means, range of question difficulty, and requirements for reporting. For example, the ideal mean of a multiple-choice test containing questions with four alternatives is 62.5%. This is the midpoint between chance selection (25%) and perfection (100%). The range of difficulty of multiple-choice questions is expected to vary from 30% to 85% to ensure that students with varying ability levels are challenged.

Each dimension of the curriculum for which results are reported must contain at least six questions if the results are to be meaningful.

2. APPROVING BLUEPRINTS

Blueprint approval establishes the overall design of the test, the exact emphases given to each category for which results are reported, and the emphases given to the different cognitive levels.

The interim blueprint is reviewed by a committee of Alberta Education personnel that represents the Curriculum Design Branch (or Language Services Branch), Regional Office consultants, and the Student Evaluation and Records Branch. This committee reviews the interim blueprint and makes recommendations to the Director of the Student Evaluation and Records Branch.

The blueprint recommended by the Alberta Education committee is then reviewed by a Test Review Committee, which consists of members nominated by the Alberta Teachers' Association, the Conference of Alberta School Superintendents, post-secondary institutions, and Alberta Education. This committee makes recommendations to the Director of the Student Evaluation and Records Branch.

3. DEVELOPING TEST QUESTIONS

Following blueprint approval, committees of practising classroom teachers working at the appropriate grade level are formed, and question development meetings are held. These committees develop new test questions that reflect the goals and objectives of the Program of Studies and curriculum specifications. Where necessary, question developers are trained in the principles of question construction. Questions built in committee are then screened for format, validity, blueprint 'fit', and other design considerations.

4. CONSTRUCTING AND ADMINISTERING FIELD TESTS

After careful editing and formatting of questions developed by the teacher committees, field tests are constructed. Any required artwork is completed during this phase of the test development process.

With permission from school and jurisdiction personnel, field tests are sent to a number of teachers throughout Alberta. The students involved are representative of the student population for which the test has been designed. A minimum sample of 150 students writes each field test.

Teachers who administer a field test are asked to comment in writing on the following:

- reading level
- how closely the question matches the way in which a concept was taught
- level of difficulty of the questions
- quality of the questions and graphics
- errors of any kind

The results from the administration of this initial round of field tests are used to validate content, to determine difficulty levels, and to ensure that questions are expressed clearly and unambiguously.

5. ANALYSING AND REVISING

The results of each field test are then analysed and scrutinized to determine whether individual questions require revision. Teacher comments regarding the way that test questions are structured and the way that a subject is being taught are also carefully considered and used to guide revision.

Questions deemed to require changes are revised and submitted for further field testing.

6. CONSTRUCTING FINAL FIELD TESTS

Once the initial field test results are thoroughly analysed and questions requiring changes are revised, final field tests are constructed. These field tests follow the approved blueprint and parallel the actual achievement test in format and design.

Final field tests, like all field tests, are submitted for further validity checking, editing, and proofreading. In grades 6 and 9, separate tests in English and in French are developed for language arts. At this point, all other tests for Grade 6 and Grade 9 are translated into French.

7. APPROVING FINAL FIELD TESTS

After the final field tests have been constructed, a second meeting of the Alberta Education Committee that represents the Curriculum Design Branch (or Language Services Branch), Regional Office consultants, and the Student Evaluation and Records Branch is convened. This committee reviews the final field tests and makes recommendations for improvement.

The Test Review Committee, which approved the blueprint in Phase Two of the test development process, meets a second time to review and recommend for approval the final field tests and the instructions for administering the tests. If a test includes short-answer or extended-writing questions, the Test Review Commitee discusses standards of achievement and marking standards appropriate for the test. Again, this committee makes recommendations to the Director of the Student Evaluation and Records Branch.

8. ADMINISTERING FINAL FIELD TESTS

The final field tests are administered and the results are used as a final screen in selecting questions for placement on the provincial achievement test. A minimum sample of 250 students writes each final field test. The sample is selected to include:

- only students who have received instruction in the course
- students representing a normal distribution of ability levels
- students from rural and urban schools
- students from large and small schools

9. CONSTRUCTING THE FINAL TEST

The construction of the final test form is based upon information collected from the final field test administration. The Test Review Committee is reconvened to review the final test form.

The test is submitted for final validity checking, editing, and proofreading. Grade 6 and Grade 9 achievement tests, in subjects other than language arts, are translated into French at this time.

For each test an information bulletin is prepared, outlining the design and nature of the upcoming tests. These bulletins are distributed to each school in September to facilitate program and instructional planning by teachers and administrators.

10. PREPARING AND ADMINISTERING THE FINAL TEST

The completed achievement test is commercially printed and prepared for distribution. It is administered to the students by their classroom teachers.

Sufficient copies of the test are mailed to each school. Quantities are based on the number of students enrolled in the subject as reported to the Student Evaluation and Records Branch by school superintendents.

11. MARKING

All written-response sections of the tests are marked by classroom teachers. These teachers, who are recommended by their superintendents, are currently teaching the course being evaluated, have taught the course for a minimum of two years, and hold a valid permanent Alberta Professional Teaching Certificate. Student Evaluation and Records Branch staff train and supervise the teachers during the marking sessions. All multiple-choice responses are machine scored.

12. ANALYSING AND REPORTING THE RESULTS

Once the test has been written, at least 20 classroom teachers review the test question by question, to judge the appropriateness of the standard built into the test. These teachers identify a test score that reflects student performance at a standard of excellence and a test score that reflects student performance at an acceptable level, based on the requirements of the Program of Studies. The teacher assessments are then compared to the actual levels of student achievement on a provincial basis. These results are reviewed by the Test Review Committee, which reconvenes a final time. This committee reviews the results of the test in terms of the objectives of the Program of Studies being measured.

A statistical report is prepared and distributed to superintendents, school principals, Alberta Education officials, and other Departments of Education. This report is also made available to the general public. In addition to the *Provincial Report*, each school and jurisdiction receives a statistical summary for its respective student population.

For further information, please refer to the respective Achievement Test Bulletins, or call the Associate Director, Achievement Test and Diagnostic Evaluation Program, at 427-2948.

APPENDIX C
STANDARD SETTING

RATIONALE

The purpose of standard setting in the Achievement Testing Program is to answer the question of whether provincewide performance is satisfactory. To use standard setting in this way requires two distinct judgments. The first is to establish what percentage of students tested can be expected to achieve at least an acceptable level of skill and knowledge required to proceed to the next level in that subject, assuming there are adequate teaching and resources. The second is to establish the test score that represents that level.

Satisfactory provincial performance can be said to occur when the percentage of students scoring at or above the established test score is equal to or greater than the expected percentage. Two similar judgments must be made for any other standards required, such as the level of skill and knowledge that reflects excellence.

Standard-setters must have a shared concept of the skills and knowledge of the borderline students for each standard set. Experience has shown that it is reasonable to judge that the expected percentage of students who should achieve the acceptable level is 85% and the expected percentage who should achieve at the level of excellence is 15%.

ESSENTIAL ELEMENTS OF THE PROCEDURE USED

1. Standard-setters were selected who were familiar with both the curriculum and the characteristics of the students who wrote the test.
2. The rationale for and the purpose of standard setting were explained.
3. The characteristics of borderline students were discussed, with emphasis on those characteristics that affect responses to the achievement test in question, and a consensus was reached for each standard set.
4. Standard-setters made and recorded judgments on a question-by-question basis for the acceptable skills standard.
5. Standard-setters made and recorded global judgments for the test at both the acceptable level and the standard of excellence.
6. The raw score derived from question-by-question judgments for each standard-setter was calculated. The standard-setters were informed of their individual standards and of the median for the standard-setters as a group.
7. The standard-setters were presented with data on the actual distribution of scores and the actual response frequencies.
8. The standard-setters were allowed to revise their judgments, but it was stressed that they need not consider the actual results.
9. The revised judgments were used to determine the test scores representing each level.

APPENDIX D
EQUATING MULTIPLE-CHOICE TEST SCORES

Comparing achievement in two different groups requires some common measure. The Student Evaluation and Records Branch develops new achievement tests for each administration, reflecting changes in curricular emphases and refinements in test design learned from earlier tests. Thus, scores from one administration of an achievement test cannot be directly compared with scores from another administration.

Various techniques are available to address this problem. In order to compare the 1989 achievement test results with results for the same subjects in 1985, the branch chose a variation of test equating as one of the techniques to be used.

Each 1985 test was administered in 1989 to a sample of students who were also writing the 1989 achievement test in that subject. The 1985 tests were administered either one week before or one week after the 1989 tests were written, using the same instructions that were used in 1985. These students had not been exposed to the 1985 tests prior to writing them in 1989.

The 1985 tests that were re-administered were scored using the same keys that were used in 1985. Scores were matched by student name and school with their scores on the 1989 achievement tests. Students for whom only one score was available were removed from the sample.

The two sets of scores in the sample were then assumed to represent the same range of achievement, as the same sample of students had produced both sets. Both tests measured achievement in the same curriculum.

Thus, achievement at a particular level in one set of scores should be equivalent to achievement at the same level in another set of scores. A score at the 20th percentile on the 1989 test would be equivalent to a score at the 20th percentile on the 1985 test, because the percentiles were based on exactly the same students. (A percentile score represents the percentage of the sample of population scoring at or below that particular score.)

For example, a score of 40 on the 1989 Grade 9 Science Achievement Test fell at the 18.7th percentile on the sample group. On the 1985 test, the 18.7th percentile came between a score of 39 (percentile rank of 17.9), and 40 (percentile rank of 19.7) in the sample group. Thus the score of 40 on the 1989 test was assigned an equivalent score of 39.4 on the 1985 test.

This fact was used to calculate equivalent scores for the two tests. The 1985 test was used as the anchor test. Each 1985 score was converted to its percentile equivalent. Percentile scores for the 1989 scores in the sample were also calculated. A transformation table was then produced.

The transformation table was used to convert all scores in the 1989 population to their equivalents in the 1985 scores. It was then possible to treat the estimated scores as equivalent to the actual scores of the 1985 population and perform a comparison using the two-tailed t-test to determine whether there were statistically significant differences between the achievement scores of students who wrote the 1985 test in 1985 and students who wrote the 1989 test in 1989.

APPENDIX E
EQUATING WRITTEN-RESPONSE TEST SCORES

The inclusion of a written-response section in the Grade 3 English Language Arts Achievement Test and the Grade 6 Social Studies Achievement Test required modification to the test-equating method described in Appendix D, page 64, for tests with multiple-choice sections only. A consideration was the possible difference in the application of the scoring system between 1985 and 1989. Scale descriptors have been revised and marker training procedures have been further refined, so scores cannot be assumed to be comparable. It was necessary to develop procedures to overcome these discrepancies.

PROCEDURES

The 1985 written-response tests were administered to the same sample that wrote the 1985 multiple-choice tests in 1989. The 1985 language arts test and the categorically scored portion of the social studies test were then scored using the 1989 scale descriptors by markers who had participated in the scoring of the 1989 written-response tests. The 1985 scoring guides were used for the analytically scored portions of the social studies test (questions 1 to 4). Samples of about 250 papers written in 1985 were re-scored at the same time by the same markers. Markers did not know the year in which any particular paper had been written.

ANALYSES

Results of the 1989 scoring of the 1985 papers written in 1985 were compared with results of the 1985 scoring for the same paper. Conversion tables were generated using the equipercentile method described in Appendix D. These tables were used to convert the written-response scores for the 1985 papers written and scored in 1989 to the same scoring standard that was applied in 1985.

Total scores for the 1985 papers written in 1989 were then calculated using the multiple-choice scores produced by using the same key that was used in 1985 and the adjusted written-response scores. These total scores were then compared with the same students' total scores on the 1989 tests in order to produce conversion tables based on the entire test. This procedure, using entire test scores while still adjusting written-response scores for changes in marking standards, gives more reliable results than separate analyses for the two parts of the tests (written response and multiple choice).

APPENDIX F
REPORTING TO PARENTS
ANSWERS TO FREQUENTLY ASKED QUESTIONS

What are the achievement tests?

The achievement tests are provincial government tests administered in Alberta schools to grades 3, 6, and 9 students in language arts, social studies, science, and mathematics.

What is the purpose of the achievement tests?

The achievement tests provide information about what students know and can do in language arts, social studies, science, and mathematics. The tests enable Alberta Education to monitor the level of achievement of students throughout Alberta. The results also help local school boards, principals, and teachers identify the strengths and weaknesses in their implementation of these subjects.

How many achievement tests will my child have to write?

Students write only one achievement test in Grade 3, one in Grade 6, and one in Grade 9. Tests are rotated so that a different subject is tested each year. In 1989, Grade 3 students wrote the language arts achievement test, Grade 6 students wrote the social studies achievement test, and Grade 9 students wrote the science achievement test. In 1990, Grade 3 students will write the mathematics achievement test, Grade 6 students will write the science achievement test, and Grade 9 students will write the language arts achievement test.

How should I prepare my child to write an achievement test?

No preparation beyond normal classroom instruction is required to write an achievement test. While students should be encouraged to do their best, a good night's sleep and a relaxed, confident approach to testing are the best possible preparation.

How much do these tests count for my child?

The achievement tests do NOT affect students' final marks. The classroom teacher is responsible for evaluating students and awarding final marks. Achievement test results are not released by Alberta Education until October, long after students' marks have been determined by the classroom teacher.

How do achievement test results help classroom teachers?

Achievement test results provide feedback on student achievement to school boards, principals, and teachers. For example, teachers in a school that consistently scores high on one part of the curriculum but low on another may wish to examine their programs to see if changes are needed to achieve a better instructional balance.

What are the limitations of the achievement tests?

Paper and pencil tests cannot easily measure such things as laboratory skills, small group discussions, and creative thinking. Thus, some student strengths can be evaluated only by the classroom teacher. Also, a single test cannot reveal as much about a student's development and growth as can evaluation by the classroom teacher over the course of a full school year.

What advantage do achievement tests have over other standardized tests?

Unlike commercially developed tests, achievement tests are based specifically on Alberta's programs of study and are designed, written, and evaluated by experienced classroom teachers from across the province. Tests developed elsewhere may not reflect curriculum or standards appropriate for Alberta.

How do I interpret achievement test results?

The Achievement Testing Program *Provincial Report* includes guidelines for interpreting results. Readers are cautioned not to overgeneralize conclusions based on a single administration of the test. Results should be compared to expectations or with the results of previous achievement tests in the same subject. Any trends that are observed in the scores must then be interpreted in the context of a variety of factors that could affect student achievement, such as the school and community environment, students' socioeconomic background, and available learning resources.

Comparisons between districts, schools, or classrooms are likely to prove misleading and are therefore discouraged.

Can I find out how my child did on the achievement test?

Individual results on the achievement tests are made available to school principals. Since the tests are designed to gather information on groups of students, not on individuals, individual results must be interpreted with caution.

Where can I get additional information about the Achievement Testing Program?

Bulletins describing the content of the coming year's achievement tests and the *Provincial Report* describing the results of the previous year's testing are distributed to schools each fall. Requests for copies of these publications or questions and comments regarding the Achievement Testing Program should be directed to:

Mr. Dennis Belyk
Associate Director
Achievement Tests and Diagnostic Unit
Student Evaluation and Records Branch
Alberta Education
Devonian Building, West Tower
11160 Jasper Avenue
Edmonton, Alberta T5K 0L2

APPENDIX G
RESULTS IN RELATION TO STANDARDS

The discussion in the main body of this report deals with results only for those students who wrote the June 1989 achievement tests. Some students were exempt from the tests, and others were absent on the day of testing. This appendix presents another way of looking at the provincial results in relation to standards. The percentage meeting standard is affected by which students are included in the population.

We are interested in the differences between the percentage of students meeting the acceptable standard when absentees and exempted students are included as part of the population compared to when they are not included.

This appendix presents the results in relation to the acceptable standard for Grade 3 English Language Arts, Grade 6 Social Studies, and Grade 9 Science. The acceptable standard was established by a standard-setting procedure (see Appendix C, page 63).

GRADE 3 ENGLISH LANGUAGE ARTS

In order to achieve the acceptable standard for the Grade 3 English Language Arts Achievement Test, students needed a combined score of 51 marks out of 100.

This part of the report gives the percentage of students achieving and not achieving the acceptable standard. Grade 3 students in Francophone or French Immersion programs are not included in the analysis, neither among those writing nor among those exempt, although many of them did write the test. Because of the reduced English Language Arts instruction in the first three grades in these programs, the standards are inappropriate for these two groups.

Figure G-1 presents the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis.

Figure G-1

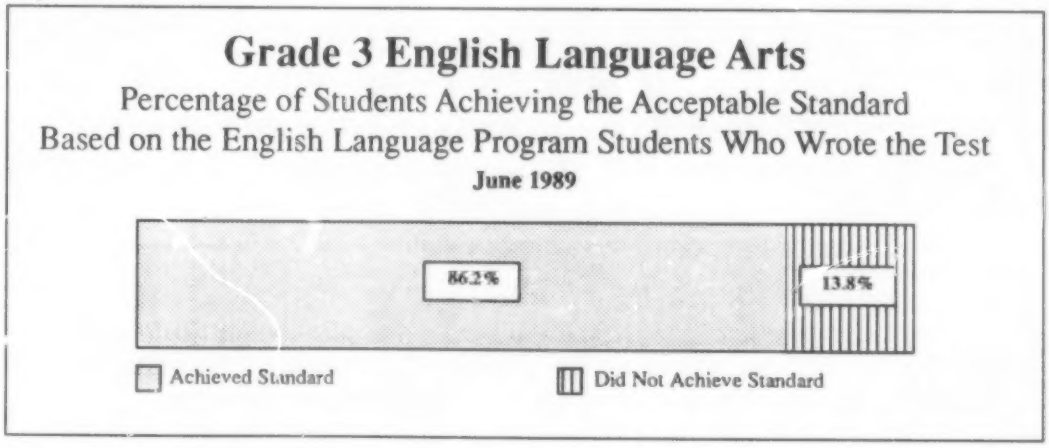
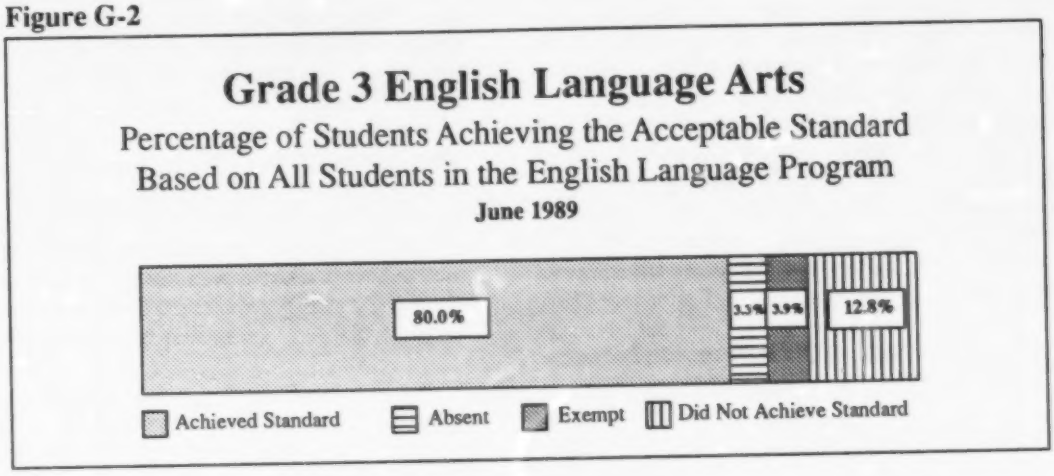

Grade 3 English Language Arts
Percentage of Students Achieving the Acceptable Standard
Based on the English Language Program Students Who Wrote the Test
June 1989
86.2%
13.8%
Achieved Standard
Did Not Achieve Standard

Figure G-2 shows the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis and those who were absent or exempt.

Figure G-2



GRADE 6 SOCIAL STUDIES

In order to achieve the acceptable standard for the Grade 6 Social Studies Achievement Test, students needed a combined score of 47 marks out of 100.

This part of the report gives the percentage of students achieving and not achieving the acceptable standard. Grade 6 students in Francophone or French Immersion programs are not included in the analysis, neither among those writing nor among those exempt, although many of them did write the test in Franch translation, and a few wrote in English. We realize that language of testing and language of instruction are important factors in student achievement.

Figure G-3 presents the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis.

Figure G-3

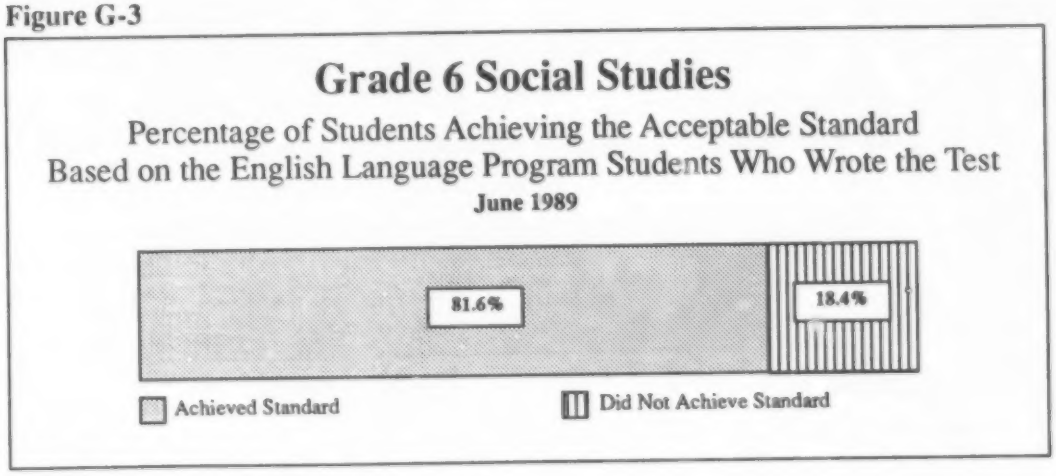
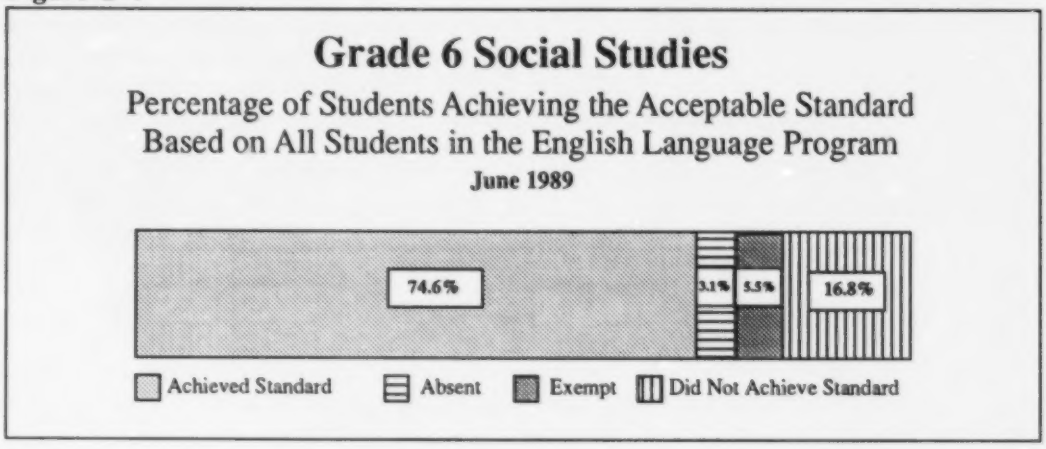

Grade 6 Social Studies
Percentage of Students Achieving the Acceptable Standard
Based on the English Language Program Students Who Wrote the Test
June 1989
81.6%
18.4%
Achieved Standard
Did Not Achieve Standard

Figure G-4 shows the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis and those who were absent or exempt.

Figure G-4



GRADE 9 SCIENCE

In order to meet the acceptable standard for the Grade 9 Science Achievement Test, students had to achieve a raw score of 38 marks out of 75.

This part of the report gives the percentage of students achieving and not achieving the acceptable standard. These results are based only on students who were enrolled in the English language program and not on the total population. Results achieved by Francophone and French Immersion students were not included in this analysis. We realize that language of instruction and language of testing are important factors in student achievement.

Figure G-5 presents the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis.

Figure G-5

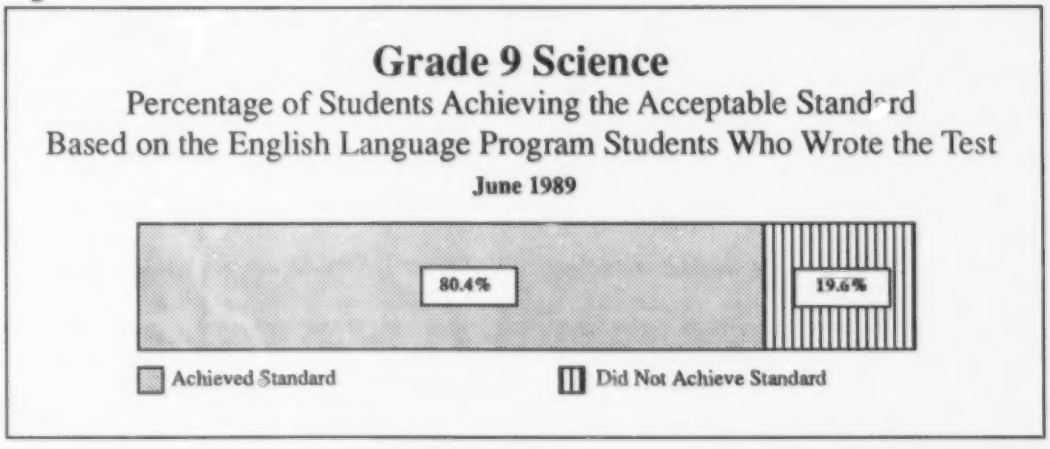
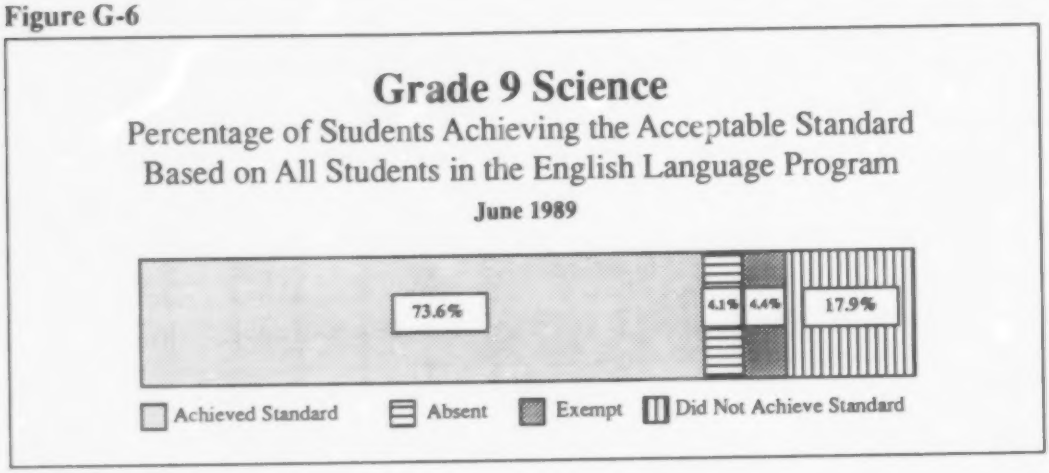


Figure G-6 shows the percentage of students achieving the acceptable standard and not achieving the acceptable standard based on those students who wrote the test and were included in the analysis and those who were absent or exempt.

Figure G-6



ACHIEVEMENT TESTING PROGRAM
PROVINCIAL REPORT
QUESTIONNAIRE

The Student Evaluation and Records Branch strives to produce documents that will be useful to the educational community. The purpose of the following questionnaire is to garner your opinions about the *Provincial Report* so that these opinions can be considered when the content and format of the report are reviewed prior to June 1990.

Please take a moment to respond to the following questions. Then detach this sheet and send it to:

Mr. Michael Robinson
Assistant Director, Data Analysis and
Student Records Services
Student Evaluation and Records Branch
Alberta Education
11160 Jasper Avenue
Edmonton, Alberta T5K 0L2

USE OF THE REPORT

1. Please check the box beside the statement that applies to you.

   My present role is primarily that of

   - teacher ☐ 1
   - school administrator ☐ 2
   - central office administrator ☐ 3
   - school board member ☐ 4
   - other (please specify)_____

2. Please check the box beside the statement that applies to you.

   I read the report, but I DID NOT use it to interpret the results attained by my students. ☐ 5

   I read the report, and I used it to interpret the results attained by the students in

   - my classroom ☐ 6
   - my school ☐ 7
   - my jurisdiction ☐ 8

3. Please respond to the following statement if you have checked boxes 6, 7, or 8 above.

   I have made use of the results to alter the educational program offered in

   - my classroom ☐ 9
   - my school ☐ 10
   - my jurisdiction ☐ 11

CONTENT OF THE REPORT

1. Please judge the usefulness of the information included in the various sections of the report by checking the appropriate boxes below.

	Very Useful	Adequate for Use	Of Some Use	Of No Use
Section 1: Summary of Achievement Test Results	☐ 12	☐ 13	☐ 14	☐ 15
Section 2: Guidelines for Interpreting Achievement Test Results	☐ 16	☐ 17	☐ 18	☐ 19
Sections 3 to 5: Specific Achievement Test Results	☐ 20	☐ 21	☐ 22	☐ 23
Section 6: Student Achievement by Gender	☐ 24	☐ 25	☐ 26	☐ 27
Section 7: Student Achievement Over Time	☐ 28	☐ 29	☐ 30	☐ 31

2. If you wish, please comment further on the content of the report in the space below.

FORMAT OF THE REPORT

1. Please judge the usefulness of the report's format by checking the appropriate boxes below.

	Very Useful	Adequate for Use	Of Some Use	Of No Use
Organization into Separate Sections	☐ 32	☐ 33	☐ 34	☐ 35
Double-Column Presentation of Text	☐ 36	☐ 37	☐ 38	☐ 39
Presentation of Figures	☐ 40	☐ 41	☐ 42	☐ 43
Presentation of Tables	☐ 44	☐ 45	☐ 46	☐ 47
Blending of Information in Text, Figures, and Tables	☐ 48	☐ 49	☐ 50	☐ 51

2. If you wish, please comment further on the format of the report in the space below.